Semi-automatic fold detection in geological outcrops using Laser Scanning data.

Applied to a limestone-marl sequences in the Vocontian Basin, France.

MSc. Thesis by Joppe Roebroeks
Department of Geoscience and Remote Sensing
Delft University of Technology
July 12, 2018

TU Delft

Semi-automatic fold detection in geological outcrops using Laser Scanning data.

Applied to a limestone-marl sequences in the Vocontian Basin situated in France.

By

Joppe Roebroeks

In partial fulfilment of the requirements for the degree of

Master of Science
in Geoscience and Remote Sensing

at the Delft University of Technology,
to be defended publicly on Thursday July 12, 2018 at 15:30 PM.

Student number:	4244214	
Thesis committee:	Dr. R. C. Lindenbergh,	TU Delft, daily supervisor
	Dr. H. A. Abels,	TU Delft
	Prof. Dr. G. Bertotti	TU Delft

Abstract

Geological surveying is a common practice performed by geologists nowadays. Determining layer orientation parameters and identifying folds are examples of features needed in order to create a geological map or model of the (sub) surface. Surveying is expensive and time consuming, therefore such survey is recently being investigated and implemented in 3D point clouds. This thesis provides a method to quickly obtain geological information such as orientation parameters and fold information in a semi-automated manner. Obtaining this parameters is currently still a difficult problem because of data density and calculation complexity.

Fast and reliable are terms important for the created method. Large point clouds can easily jam a system and reliability is important for usability and compatibility of the created results. A method is designed which can handle bulk data, performing operations such as filtering and noise reduction assisting in data handling. Next to general calculations such as roughness filtering, intensity filtering and directional filtering of generated normals, the use of a Graphical User Interface (GUI) greatly improves the reliability of the feature calculations. The user selects areas which are potentially interesting for acquiring layer orientation information, which are then analyzed by the method itself creating a base for determining the layer orientation parameters. This base is then used to attempt the reconstruction of the bedding layers of the geological outcrop resulting in the desired geological features.

The method has been tested on the “La Charce” dataset and the “Meso-scaled fold” dataset for method validation and on the “Pradelle 2” dataset for compatibility checking. Average calculation time is with approximately 10 minutes reasonable fast. The acquired dip direction of approximately 156 degrees is very similar to that of Prentice (2017), Bisschop (2017) and the fieldwork teams of 2016. The dip angle of 52 degrees is on the other hand off compared to validation results. The layer thickness calculations are comparable, all theses show that the bulk thicknesses between ~30 and ~70 cm. The fold dataset has no validation results but the images look good. It has to be mentioned that all results obtain commission and omission errors but these are in no comparison to the results which are correct.

Acknowledgements

I would like to thank Roderik Lindenbergh, Hemmo Abels and Giovanni bertotti for supervising me and helping me during these 9 months of work. With their help, time, feedback and support, this thesis was a very pleasurable experience resulting in a research which will be the base for further development.

This research was based on work provided by Elizabeth Prentice, whom I would therefore also like to thank. Without her work it would not be possible to make a script in the first place. I would also like to thank her for her valuable insight in her own made script. A special thank you to Amerigo Corradetti for providing me with a dataset on which I could test my fold identification method.

Lastly I would like to thank my family and girlfriend for providing mental support and a listening ear for my concerns and problems.

Contents

1 Introduction

Rock outcrops have been studied extensively in the field of geology. They are of great importance in the understanding of the geological history of a region. This history includes the composition of the rock (and thus the evolution of the environment) and the tectonic history. This research aims to contribute to this understanding by analyzing rock outcrops in 3D. The research is a between the department of Geoscience and Remote Sensing and Applied Earth Sciences. There are several advantages of analyzing outcrops in the 3D environment. Firstly, whole outcrops can be evaluated instead of just physically accessible parts of the outcrops. Secondly, virtually analyzing the outcrop may lead to information which may not be obvious on first sight. Lastly, when digital processing of these outcrops is viable, vast areas can be evaluated within a reasonably short time.

The digital representation of vast areas in a digital environment is often done by point clouds. Point clouds provide a good representation of the reality preserving units and distances in space. The disadvantage of point clouds is that reality is sampled and thus details are lost. However, the density and accuracy is often high enough to recreate reality up to a certain extend. One of the problems with point cloud processing is the amount of data. Often there is too much data to be processed, this is therefore then also one of the challenges faced nowadays; more efficient and smart processing is needed in order to process all available point cloud data.

Figure 1: Bedding planes and fold [6]

In this research, bedding planes and folds in rock outcrops will be extracted by means of point cloud processing. In addition to that, some statistical analysis will be performed on those extracted beddings in order to obtain additional features which will be elaborated on in chapter 2. Figure 1 illustrates the

mentioned features, the distinctive surfaces of the layers are called the bedding layers and the bend in the layers is called the fold. Next to extracting these features alone, another challenge is to process all available data instead of a few percent. All these computations should be done within reasonable time with a standard computer.

The initial study region will be the same as that of Prentice (2017), where an outcrop near the small village called “La Charce” is being processed. This dataset has been acquired during the yearly Applied Earth Sciences BSc Fieldwork. In addition to this dataset, a dataset belonging to the Nothern Appenines in Italy is being used acquired by Corradetty (2017). The simplest dataset is that belonging to the La Charce region in France. This is a structure with very typical layers which have roughly the same thickness, this regular pattern is shown in figure 2. One can see that the originally horizontally deposited layers are now situated under a considerable angle. Some example red lines are drawn to indicate the bedding of the geological layers.

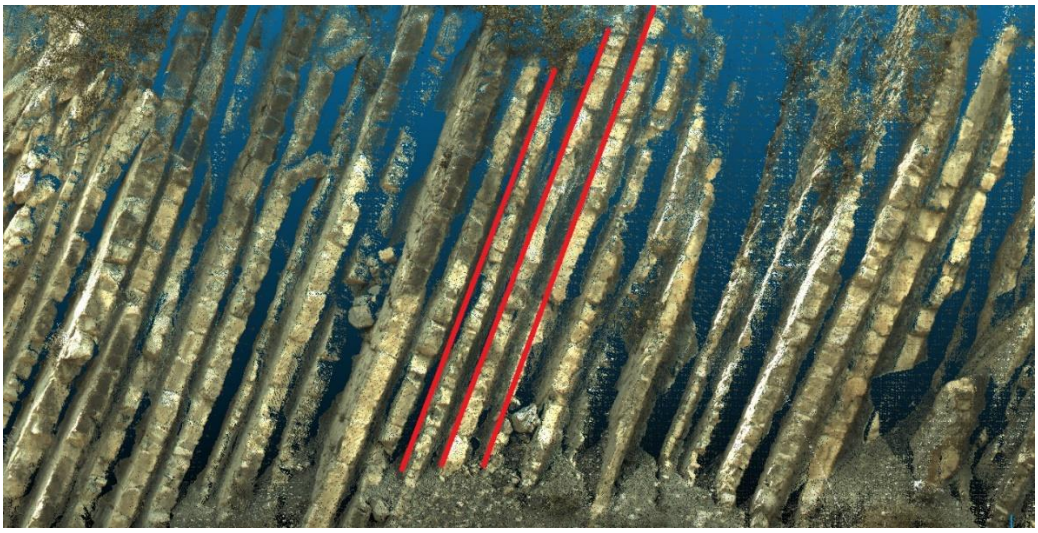


Figure 2:Point cloud data of the La Charce formation southeast France by Prentice (2017)

Some calculations for the digital analysis of rock outcrops have already been provided by Prentice (2017). Her study will be used and improved with the goals set below, at the same time answering the primary and secondary questions.

1.1. Goals

This project has three primary goals. The first goal is the extension of the methodology with fault and fold recognition in 3D point cloud data, the second goal is to make data handling more efficient and the last goal is to make the processing more autonomous. These goals result in the research questions mentioned on the next page. The goals are mentioned below.

1. *Extraction capabilities.*

    Current processing methods provide tools to extract orientation parameters (strike and dip direction) of rock layers (Prentice 2016). The goal is to incorporate more geological feature recognition algorithms such as fold detection. Even though linking outcrops is probably hard to

implement within the given time, a foundation for such results can be made. Also, looking at thicknesses of layers, analyzing thickness changes over distance and orientation changes can be important. The last mentioned properties provide information about the Milankovitch cycle and the location of the outcrop in the sedimentary basin (where was it in the past and what happened in terms of pressure?). The last goal is to provide a result which is comprehensible and summarizing for the whole process. Results need to be on point and they do not contain unnecessary information.

2. *Improve data handling capabilities.*

Point clouds processing requires a lot of computing power. Processing units generally do not have enough power to process tens to hundreds of millions of points with the same method without requiring a lot of memory and time. A smarter way has to be found to evaluate and process huge amounts of data without having to compute for days. The goal is to evaluate all points (not a partition) within hours. Fast with high quality is a priority.

3. Semi-*autonomous processing.*

The previously generated script requires various input parameters in order to properly operate. Processing should become more autonomous in order to prevent human errors. Largely autonomous processing is an important goal within this project.

1.2. Research questions

From the variety of goals and statements mentioned in the previous pages, a set of research questions have been formulated. The research questions, which are the main driver for this research, are formulated in the paragraph below.

Primary research question

To what extent can bedding layers and folds in geological outcrops be automatically detected in large 3D datasets?

Secondary research questions

1. What geometric parameters should be extracted from point clouds?
2. What are the characteristics of laser scanning point clouds?
3. What existing point cloud processing methods could be useful for such feature extraction?
4. How does a successful method look like?
5. How to efficiently extract the required information from large datasets?
6. What parameters influence the detectability of the faults/folds and to what extend?
7. How to validate the estimated parameters computed by the computer?
8. How can this method help geologists in assessing geological outcrops?

2 Digital outcrop representation

The area of interest in this research are sedimentary rocks. Sedimentary rocks are rocks which are created during the process of deposition and subsequent cementation of material that is situated at the Earth's surface within bodies of water. Sedimentation is a settling process of minerals or organic material which has been created by weathering or erosion. The resulting sedimentary rocks cover 73% of the Earth's dry surface [10] and because of tectonic activity these sedimentary rocks can be found in various shapes and conditions.

This chapter will describe the features which are going to be extracted together with the acquisition methods which are available and usable to do this. Then there will be a small elaboration on the characteristics of point clouds and the chapter will be closed with existing work on point cloud processing.

2.1. Geological features

The following subchapters include the geological features and general terms evaluated during this thesis.

2.1.1. Stratification

Sedimentation is gravity driven and therefore the original layering of sedimentary rock is more or less horizontal. This horizontal layering is visualized in the following image.

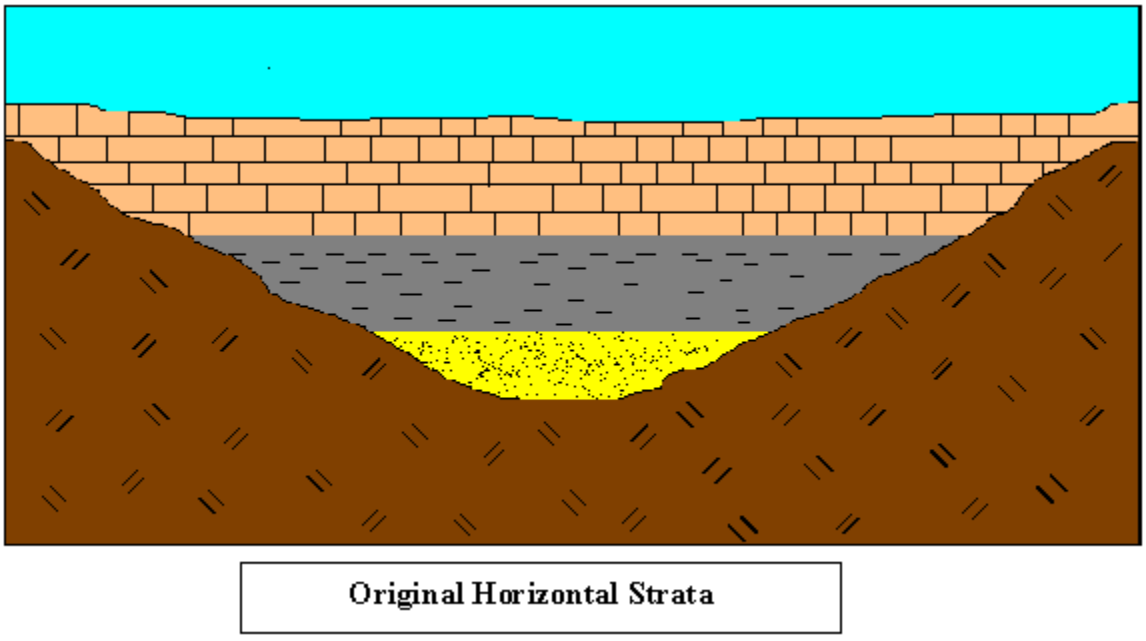


Fig 3: Original horizontal sedimentary layer [1].

Figure 3 is a simple example of stratification. Stratification is the way sediments are layered on top of each other. The stratification scale ranges from submillimeter scale to hundreds of meters or even kilometers in the horizontal direction and the layers resulting from stratification can be millimeters to several meters thick. Vertical stratification is a fundamental feature of sedimentary rocks.

The most common and very large scale of stratification is shown below in figure 4a, the sedimentation takes place in a marine environment where sandstone and limestone is formed on the sea floor through transportation of water containing small particles of sand or mud.

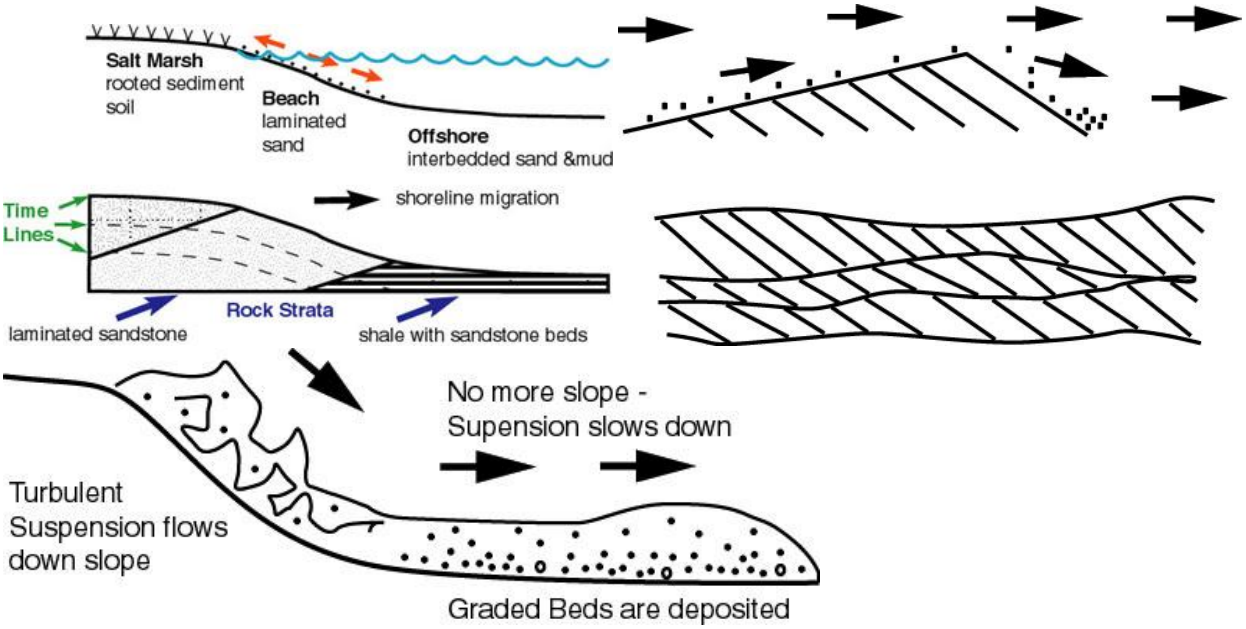


Figure 4a (left top): Marine sedimentation, Figure 4b (right top): Cross-bedding, Figure 4c (left bottom): Graded bedding. [2]

Figure 4b shows a strata which is often seen in sand areas such as deserts but is also visible on the sea floor. Wind (or water) transports sand particles in the direction of the flow (indicated with arrows) and deposits the particles when the speed drops and the particles can no longer be carried by the medium. Sand particles can be eroded from the same layer as it is being deposited in. However, if there is more deposition than erosion, cross-bedded sandstone layers such as shown in figure 4b can be found. The layers are inclined in the transportation direction and flow, the layers can be used to estimate wind speed velocities and to determine the direction of paleoflow. There is also a type of bedding which is commonly associated with turbidity currents described by Brittanica (2014). The slope between a continental shelf and a deep sea basin is often the playground for turbidity currents. Over time sediment is being deposited on this slope, a high energy event such as an earth quake can trigger these sediments to move down the slope. These moving sediments mix with water and (often) drag more sediment with it as it moves down the slope creating a turbid layer of higher density (suspended sediment) water. As soon as the layer of moving water reaches the bottom of the slope, the sediments start to deposit again resulting in a graded bedding showing a grade in grain sizes. Particles are being deposited as the suspension slows down, showing a large grain size which is decreasing over distance in the direction of flow. The newly formed bed is often covered with small sized particles coming from the final sedimentation of the “dust cloud” formed by the movement, making it difficult to identify the graded beds on the surface.

2.1.2. Milankovitch cycles

Repetitive packages of sedimentary rocks are often observed in nature (stratigraphic cycles), these repetitive packages follow from local climate regimes which change over thousands of years as

mentioned by E. (2008). One of the factors influencing the climate (and thus the repetition) is the Milankovitch Cycles. The Milankovitch cycles describe the way the earth is circumnavigating the Sun and therefore determines how much solar radiation The Earth is receiving at any given location and time.

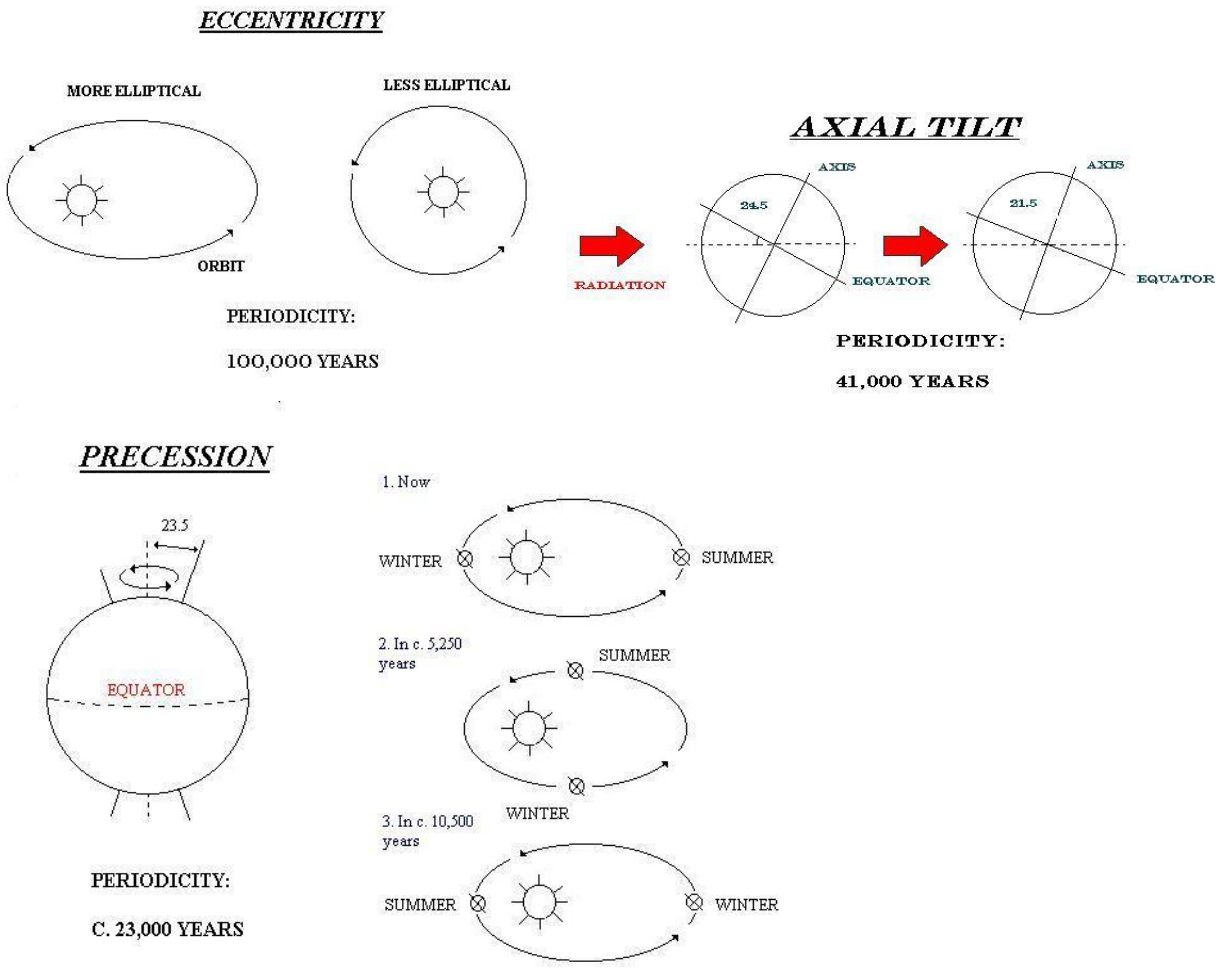


Figure 5a: Eccentricity, Figure 5b: Axial Tilt, Figure 5c: Precession. [3]

Figure 5 show the three different Milankovitch cycles possible. The first one shown in figure 5a is the eccentricity (or ellipticity) of the Earth's orbit around the sun, the more elliptical the orbit, the larger will the deviations in the received sunlight will be. The Earth's ellipticity fluctuates between 0 to 5% with a cycle of 100,000 years [3]. Figure 5b shows the axial tilt of the Earth with respect to the plane of orbit around the sun. The axial tilt fluctuates between 21.5 degrees and 24.5 degrees, and one hypothetical influence of this Milankovitch cycle on the climate is that with a smaller axial tilt the growth of ice sheets is stimulated. Figure 5c shows the precession of the Earth's rotation, this can be explained as a wobble of the rotation axis as the Earth is spinning. This wobble is similar to that of a whirlabout when it's decreasing speed, the only difference is that the periodicity of the Earths wobble is 23,000 years.

All of these three Milankovitch cycles bring changes in the climate of The Earth as mentioned by Grippo (2014), resulting in huge changes in local regimes anywhere on the surface of The Earth. Therefore it's

reasoned that these cycles are one of the major causes for the repetitive packages of sedimentary layers all around the world.

2.1.3. Geological features

Tectonic activity bends and breaks the rocks resulting in various interesting structures. Three of the major important large scale changes to sedimentary rocks resulting from tectonic activity or significant rock displacements are changes in orientation and the creation of folds and faults. Those three important features are shown in figure 6.

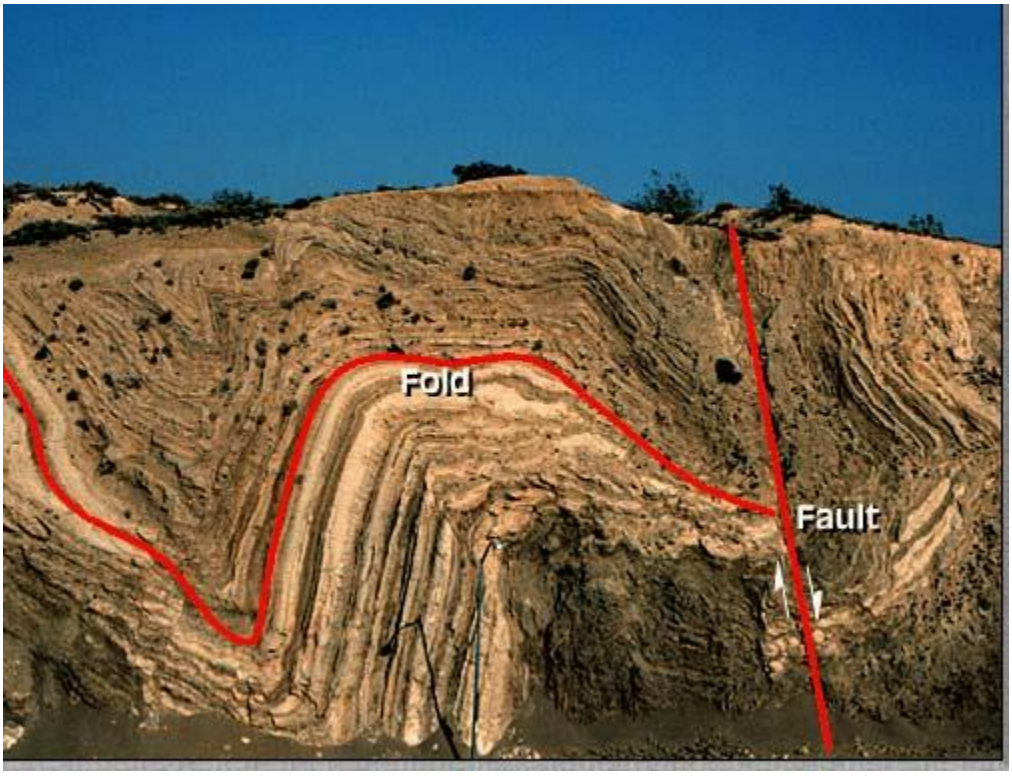


Fig 6: folds, faults and orientation [13].

In figure 6, both folds and faults have been indicated with red. Folds are created when horizontal pressure deforms the layer causing it to bend in a certain way without causing too much deformation or going too fast such that the layer breaks. When there is too much pressure applied on the layer and the deformation is too much or the deformation process takes place too fast, the layer breaks and slides resulting in a fault.

Layer Orientation parameters are characteristics which determine the orientation of the layers. The parameters are the dip direction and dip angle. The dip direction is the direction the layers are tilting towards which is quantified with an angle between 0 and 360 degrees. The dip angle is the angle of the layers with respect to the horizon quantified by an angle between 0 and 90 degrees. These orientation parameters result from large pressure-based changes which tilt layers (as you can also see in figure 6). These layers are therefore displaced from their original horizontal position and now dip towards a certain direction (dip direction) under a certain angle (dip angle).

Figure 6 contains most features which will be extracted from 3D datasets processed during this thesis. The first interesting formation to investigate is the geological layering as is it found in their most simple form. The other main features which will be investigated are with respect to layer thickness, orientation parameters and the identification of folds. For every feature, a set of parameters with their own characterizations can be identified, answering the following question:

- **What geometric parameters should be extracted from point clouds?**

A list of every feature or also called geometric parameter is given below in short and after that elaborated.

- Layer orientation parameters
- Potential curvature in layers
- Layer thickness
- Orientation of layers on either side of a fold
- Axial surface

These features are not necessarily geological features but features which are interesting to investigate when considering geological outcrops. These parameters will be described in more detail below.

Layer orientation parameters are the first parameters which are identified by any geologist. They are called the dip direction and dip angle and are defined by the direction and angle under which the layers are situated. The dip direction is the angle the layers are dipping towards with respect to the magnetic north (more or less similar to the geographical north). The dip angle is the angle which the layers make with respect to the horizon. Thus, the layer orientation parameters are two angles, one is an angle with respect to north and the second angle is with respect to the horizon.

Sometimes, **curvature** can be present in layers. Extensive pressure on geological layers can cause layers to bend slowly. This curvature is not always observable with the naked eye and that is why digital analytics comes in handy, the slight curvature can be obtained by looking at the distance from all points to a perfectly flat fitted plane. If a curvature is present, a distance pattern will occur when plotting these. The top and bottom of the point cloud (belonging to the plane) should show a distance to the plane which is opposite to the middle of the point cloud. If the top and bottom of the point cloud are above the fitted plane, the middle of the point cloud should be below the fitted plane and vice versa. The curvature in 1 layer is defined by a difference in angle in the local planes fitted within the bigger layer.

The **thickness of layers** is defined as the perpendicular distance between the bedding of two consecutive geological layers in meters. The distance is most conveniently calculated along a 3 dimensional line. This line crosses all fitted planes at a fixed height at the middle of the total vertical interval. This 3D line will also be evaluated later in chapter 3.4.9.

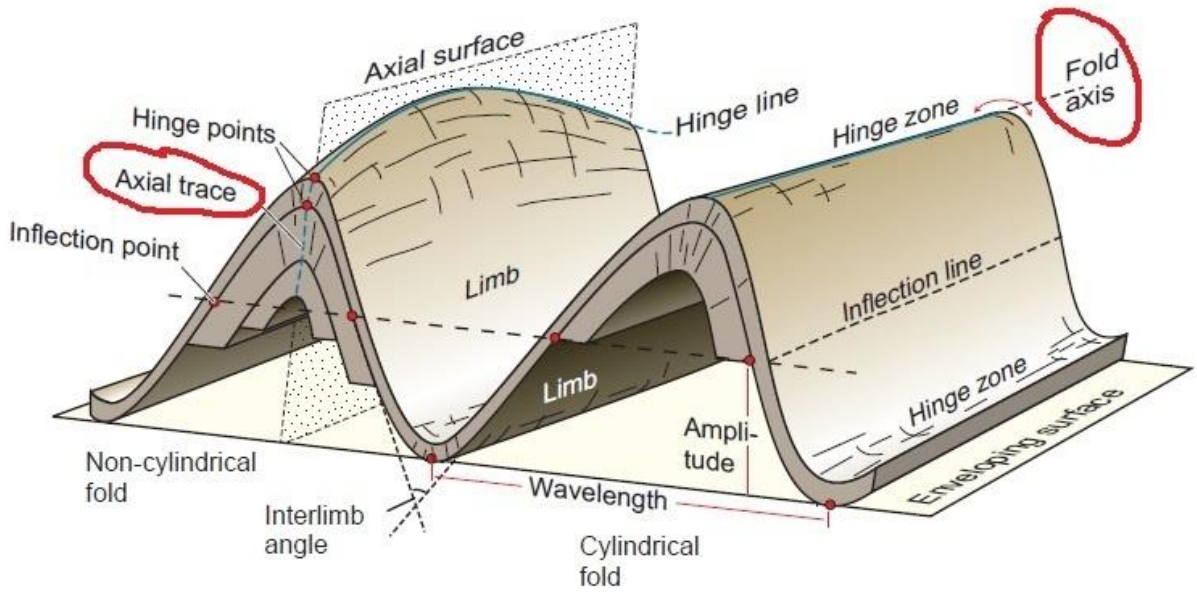


Figure 7: Axial surface and fold axis.

Folds have different **orientation parameters on either sides of the fold**, these orientation parameters are required to derive the axial surface. They are defined in the same way as the earlier mentioned "Layer orientation parameters", but now have to be defined twice for two sets of layers on either sides of a fold.

The **axial surface** is shown in figure 7. The axial surface is a plane in 3D space which defines the border between the two limbs of a fold. On either side of the axial surface the layers are dipping with a different angle and into a different direction. The axial surface is defined by a simple 3-dimensional plane with a position in space.

2.2. Digital outcrop sampling

In recent years, optical devices enabled the digital sampling of outcrops by means of dense 3D point clouds. One of the most time consuming tasks of is data acquisition itself. It often requires more than 50% of the total time of a project from beginning to a digital result. There are several ways of acquiring digital information on real objects. The acquisition methods considered good enough for this thesis are those with a very high point density. Acquisition methods with a very high point density are LIDAR and Photogrammetry. These methods generate a 3D environment with point densities up to 10's of points per square centimeter. Chapter 2.2 contains information about the basic principle of LiDAR, photogrammetry and GNSS, they will be described in the following subchapters.

2.2.1. Photogrammetry

The word photogrammetry originally means photo drawing measurement. Photogrammetry is the art of creating a map or 3D model of real-world object from photo's.

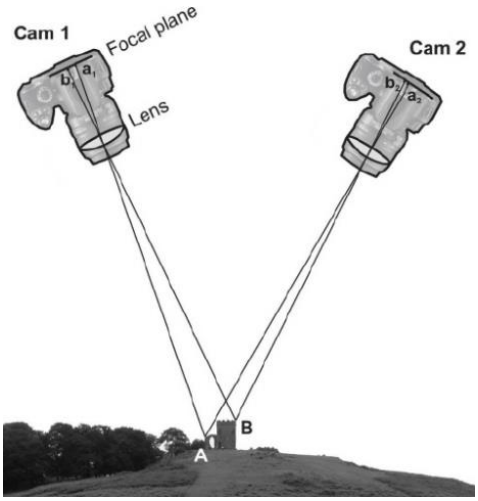


Figure 8: photogrammetry principle, found in Walstra (2007)

Figure 8 visualizes the principle of photogrammetry. Photogrammetry uses images from cameras in order to create a 3D image of an object or environment. Often tens to hundreds of images of the same object have to be taken in order to get a full digital coverage of an object.

The 3D model of the object or environment is generated by combining pictures. Pictures are combined by finding common points in several images. Because the camera is looking to an object from a different angle, the object is observed from a different perspective. The difference in perspective provides information about the position of the object in the 3-dimenstional environment.

Objects can be put in a 3D space when the object is seen in two images, and if one common object distance is known in both images. This because the object has to be referenced with respect to one reference length.

Objects are then stitched together in 3D space by means camera orientation parameters, matching, common points observed in several images and a known length (scale) observed in the images which are being stitched together. The exterior orientation parameters define the location (X,Y,Z) and viewing direction (ω, ϕ, κ) which are the rotation about the X,Y and Z axis respectively.

Many maps which are made nowadays, are made with photogrammetry. Aerial photogrammetry provides those detailed maps by using a camera which is vertically mounted on an airplane [4]. By flying over an area making pictures every few seconds, pictures are being generated with usually around 30% overlap. Together with ground control points of which the location is known, a detailed map of an environment can be made. The second large application for photogrammetry is terrestrial and Close-range Photogrammetry. Hand held cameras are being mounted on a tripod or pole facing the object. By measuring the exterior and interior orientation of the camera together with some matching points in both images, a reasonably correct 3D model of an object can be made.

Photogrammetry is used in many fields of expertise, some examples of the application of photogrammetry are building measurements, engineering structures, forensic research, imaging mines or earth-works, imaging archaeological artifacts and many more. Photogrammetry can be used in many different fields, and geological outcrop sampling is one of them.

2.2.2. LIDAR

LIDAR stands for Light Detection And Ranging or Laser Imaging Detection And Ranging. It is a remote sensing method which uses a pulsed or modulated laser signal to measure ranges to an object or to the Earth. The image below shows the principle of laser scanning.

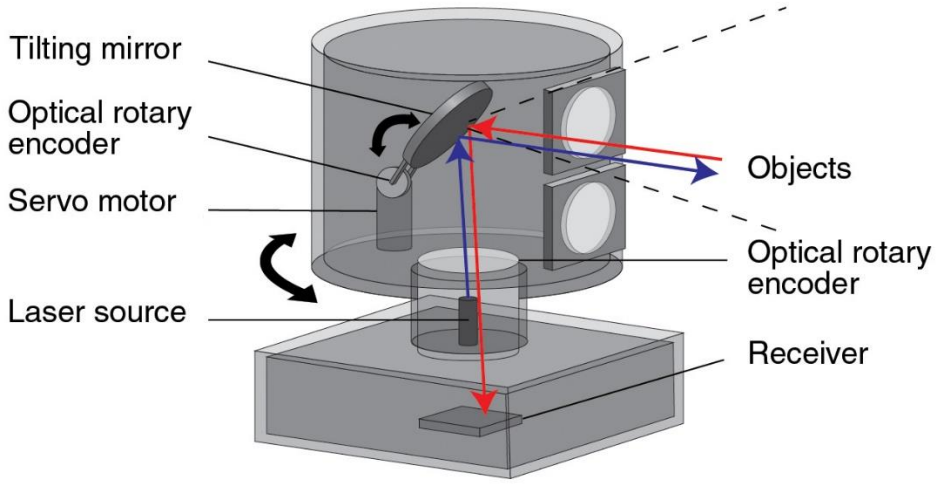


Figure 9: LiDAR principle [5]

There are 4 major components in a laser scanner [5], they listed below with their specific functions.
- Laser source: This is the emitter of the pulsed laser. In most cases this is a class 3R laser which is considered safe if handled with care. The class 3R laser often uses visible light for safety reasons. The laser provides the primary signal for measuring the distance to an object or point.
- Tilting mirror: This component provides the possibility to point the static laser source in different directions.
- Optical encoders: There are often two encoders present in a device, one encoder is used for the mirror to determine the position of the mirror. The second optical rotary encoder is used for determining the position of the laser scanning head, measurements of the positioning of the optical head are also required for precise determination of the returned signal.
- Receiver: The receiver is used for precise timing of the return signal. The origin of the signal is known because of the optical encoders, the receiver will now only determine the distance to the object. The time in between sending and receiving the signal determines the distance to the object.

The result of the use of a LIDAR instrument is a point cloud (Vosselman 2010). The resolution with which the surrounding is being sampled can often be determined by the user. The result can be a point cloud with a point density up to tens of points per square centimeter (depending also on the range to the object). LIDAR is based on true 3-D measurements and is not based on stitching as is the case with photogrammetry. Even though both methods are principally different, point cloud densities and accuracies are not necessarily differing in the same way as described by Westoby (2012).

LIDAR has a very wide range of applications. LIDAR is used on both airborne platforms and ground platforms and it performs very good on both. The airborne platform provides a fast coverage of a large area, making it viable to analyze 10's is square kilometers within hours. The major restriction of the airborne platform is the footprint of the laser, the footprint becomes larger the higher the airplane flies. The larger the footprint the more inaccurate the estimated height becomes since it is an average over a larger area. Ground-based LIDAR can provide a very detailed model of an object/environment on close range. Ground-based LIDAR is often used for indoor applications or for close range applications such as measuring building sites or streets since the range of a laser scanner is often limited to a few hundred meters.

LiDAR is also used on a UAV platform such as drones as described by the US Department of Commerce (2018). In order to use drones for such measurements, a good gyro stabilizer, IMU and Flight controller are necessary (these are necessary for all non-static acquisitions). A laser scanner can be mounted on the drone and together with the 3 mentioned components a drone can make a fairly accurate 3D point cloud of the environment. Drones can also be used for photogrammetry, the only difference is the acquisition device (which is a camera when performing photogrammetry).

2.2.3. GNSS

Point clouds and photogrammetry acquisition methods often result in data incorporated in a local coordinate system. Without information on the location and orientation of the data, the obtained data could be located anywhere in the real world. Obtaining real world positions (in a certain reference system) is performed by incorporating GNSS with a coordinate reference system such as WGS 84. With help of GNSS, the local X Y Z coordinates of the Laser scanning data (or photogrammetry data) can be transformed in latitude and longitude coordinates. Especially in the field of Geological surveying, real world coordinates are very important since orientation parameters can only be deducted when the data is properly georeferenced.

The principle of GNSS (Global Navigation Satellite System) is based on clocks and known position of the included satellites. GNSS constantly transmit information with respect to their current time and position. The current location of a satellite is determined with respect to an earth model, and the time is produced by atomic clocks present in the satellites. A GNSS receiver observes multiple satellites at the same time, receiving the position and transmission time of multiple satellites. With these positions and times, the GNSS receiver can then solve equations to acquire its own location, based on the received locations and the shift in time between the received time signal and the receivers own time signal.

2.3. Characteristics of point clouds.

Point clouds are merely tools to represent reality up to a certain extend. This means that there is a limit on how good reality can be represented. This poses certain advantages, disadvantages and challenges with respect to using and handling point clouds which result in certain characteristics of point clouds. This chapter will elaborate on those characteristics and therefore answer the following research question:

- **What are the characteristics of laser scanning point clouds?**

The first property of point clouds is the most obvious one, reality is sampled by a number of points in a 3D environment. Surfaces, corners and edges belonging to tables, walls and doors cannot be fully captured and parametrized by a remote sensing device. They are merely sampled by a number of points,

these points can be captured in color if the scanner includes a camera. As mentioned these number of points have their position in a 3D space on a 1:1 scale.

Apart from a camera there is no real additional information added to these 3D sampled representations, and thus creating objects has to be done by computing them. This leads us to the second and third characteristic of point clouds: Programs and coordinate systems. These obtained point clouds can be very large (up to 10's of millions of points) and require dedicated software in order to visualize or process them. Several types of software are available depending on the desired application, but depending on your application and budget your choice is limited because of license restrictions. The coordinate system is as important as the point cloud itself because a point cloud is largely useless if you don't know where it is located with which orientation and vice versa (of course). Most point cloud acquisition systems do not result in point clouds which are georeferenced. Proper geo-referencing is required if a point cloud is to be used for a real world application. For example this thesis, if the position and orientation of the point cloud made of a geological outcrop would be unknown, performing analysis would be useless since it cannot be compared to the real world.

Laser scanner often observe anything it can see, which means that also objects which are of no interest will be observed (if they are in line of sight). The fourth characteristic of point clouds is that they often need to be clipped in order to only maintain the information of interest. Objects which are of no interest unfortunately need to be filtered out by hand (called clipping) by "cutting" the data with a dedicated software program. Clipping is required in order to increase processing efficiency. Clipping can reduce the amount of processed points by 50% to 80% or sometimes even more. It has been mentioned that dedicated software programs are required in order to acquire the desired results. Unfortunately software programs for the specific case of this thesis are not available and are therefore also a reason for this thesis to exist. Point cloud processing of geological outcrops is still in its pubertal age, requiring much more knowledge, trial and error before it can be used on a larger scale.

General properties obtained by a laser scanner are X, Y and Z values with respect to the position of a point. Secondly, often also RGB values are obtained by using an internal camera. Lastly, the intensity of the return signal is stored, which can also be used for later computations. Intensity values are (often) only stored in case a laser scanner is used. When using photogrammetry to obtain point clouds (which is also often done), these intensity values are evidently not available. In case photogrammetry is used, synthetic intensity values can be created by converting the RGB values with certain weights attached to every color value.

Point clouds are very difficult to process because of the sheer size of the point clouds and the lack of intelligence of software. Humans have the ability to interpret point clouds and geological outcrops in a matter of seconds, since the desired properties are known. Software on the other hand, has much more data to handle, requiring much more time and effort because of the millions of points which need to be processed. The disadvantage of the sheer size is also its advantage, the result of any computation is often statistically supported by millions of calculations making the result much more robust than a human regularly achieves. It has to be mentioned that the increased level of robustness is only achieved if the method is adequately good. The result is only as good as its method, a faulty method results in faulty data.

2.4. Existing work on point cloud processing

There have been various attempts to work with Point cloud data samples of geological rock outcrops. Analysis is done both with external software and by scripts and programs created professionally or personally. There are 3 examples of free programs which are easily accessible which can be used for at least a part of such computation, they are listed below. The list includes the specific purpose of the program.

- **Lime.**
  This is a semi-free program which can be used in order to analyze virtualized outcrops, it has been used by T. Bisschop (2017). One can only obtain a license once (80 days) after which a new license needs to be acquired from the developers. With this program one can obtain geological layers and their extent.
- **Cloudcompare.**
  This is a completely free program which includes features such as normal computation, different types of segmentation, volume calculations, distance calculations, alignment of point clouds and even fracture detection. The problem for using it in a thesis such as these is that it is difficult to combine the individual processing tools to create the desired result.
  Cloudcompare is also slow when datasets become large. A large amount of calculations can only be done on 500.000 points or less (the acquired and used datasets are approximately 18-20 million points). Apart from the inefficiency, the program also misses features which are needed to come to the desired result (orientation parameters, fold identification and fault identification). Cloudcompare can however be used to generate an evenly spaced point cloud (subsampling).
- **Meshlab.**
  Meshlab has a lot of incorporated features just like Cloudcompare. However, Meshlab's purpose is even further off from the goal of this thesis than that of Cloudcompare. Meshlab includes many different tools for processing 3D pointclouds, but it has no tools for working with geological outcrops. Meshlab does include great capabilities for creating meshes from point clouds which is a great advantage for 3D surface reconstruction.

There are also commercial software programs which can do certain tasks implemented in this thesis. These software programs however, are very expensive and inaccessible. Some software packages make use of polar plots. Polar plots provide statistical insight in normal orientations of points, for every point a normal can be computed resulting in a direction wrt. north and an angle wrt. the horizon. Those angles are shown in a circular plot with the circular steps being the angle with respect to north and the radius the vertical angle. This 2D representation of 3D data is extremely powerful for initial point cloud analytics. Unfortunately, the used program (Matlab) doesn't have those features nor a point cloud toolbox and thus alternative analysis methods need to be developed.

There have been multiple persons who have been analyzing geological outcrops in 3D data. The work performed by these people together with the available processing methods mentioned in chapter 2.4 form the major part of the answer to the following question:

- **What existing point cloud processing methods could be useful for such feature extraction?**

Matasci et al (2015) and Humair F. (2015) describe several methods for analyzing geological outcrops using Terrestrial Laser Scanning point clouds.
For identifying layers in the dataset, a number of actions were performed:

- Intensity correction: The retrieved intensity diminishes over distance and with increasing incidence angle (angle with respect to the normal of the illuminated area).
- Denoising: The Nearest Neighbor algorithm of Matlab has been used in order to get rid of points with an intensity greater than the 50 nearest neighbors.
- After manual recognition of the spatial distribution of the rocks, a corrected signal strength interval has to be selected for every rock type. The intensities outside this interval will be removed from the dataset. This step is a filtering step.
- The second filtering step is based on a local density of points. A ratio of the intensity of points within two radii R1 and R2 determines if a point at the center is accepted in the next step.
- Eventually Clustering is performed by using the Label-Connected-Component tool in Cloudcompare.

The result is a segmented dataset based on intensity values. The method is able to identify different types of rocks and not necessarily different individual layers. The identification of different types of layers has not been tested with this method

In the same article, the commonly used method for detecting a fold axis was also mentioned, the Pi diagram technique. This technique uses a polar plot to identify orientation clusters of the normals of points, and according to these orientation clusters a fold axis can be determined.

In K. Anders (2016) it is described how planes and surfaces can be detected in a TLS point cloud. Here planes are automatically fitted by implementing a region growing algorithm using points with a very low curvature. This lowest curvature is always with respect to the neighbors of the points in question. The region growing of a plane continues until no more points meeting the curvature requirements, can be fitted. The size and validity of the resulting planes is very much dependent on the used parameters, which makes it a very sensitive method.

In S.J. Buckley (2009), A. Abellan (2016), R.M. Silva (2015) and Schober (2011), the emphasis was more on acquiring the data and visualization of the point clouds rather than the information contained in the dataset.

It is clear from the shown examples that modelling and processing of geological outcrops in the 3D environment is still in its early stage of development, this thesis is going to add to this digital processing environment in an interactive way.

Conclusions

There are geological features which are of interest to this thesis like bedding layers orientation and thickness or fold identification. These geological features can be captured using conventional acquisition methods such as laser scanning or photogrammetry and with help of GNSS. There are people who've been working on point cloud processing and there are programs which sole purpose is to process point clouds. Even though point cloud processing solutions are available, processing is still time consuming and computationally heavy. Therefore, new software solutions are needed in order to cope with point clouds in a better and faster way.

3 Methods

This section contains the practical approach to the research questions mentioned in chapter 1. This chapter starts with the proposed processing method divided up into the parts which describe the processing in case of just bedding identification and in the case of fold identification. After that, all individual features are elaborated on and extensively described. Then follows information on the used software and programs and a chapter with the Data requirements for point clouds. The chapter is closed with a conclusion.

3.1. Proposed processing flow

Chapter 2.4 describes some processing tools which are identified and which can be used. These are not all tools which will be used, this chapter will use some of those tools and a lot more. The order in which the processing tools are going to be used, is described in this chapter. The amount of processing steps and tools are more extensive, since new steps are necessary to improve the result and create the results faster. This chapter answers the following research questions:

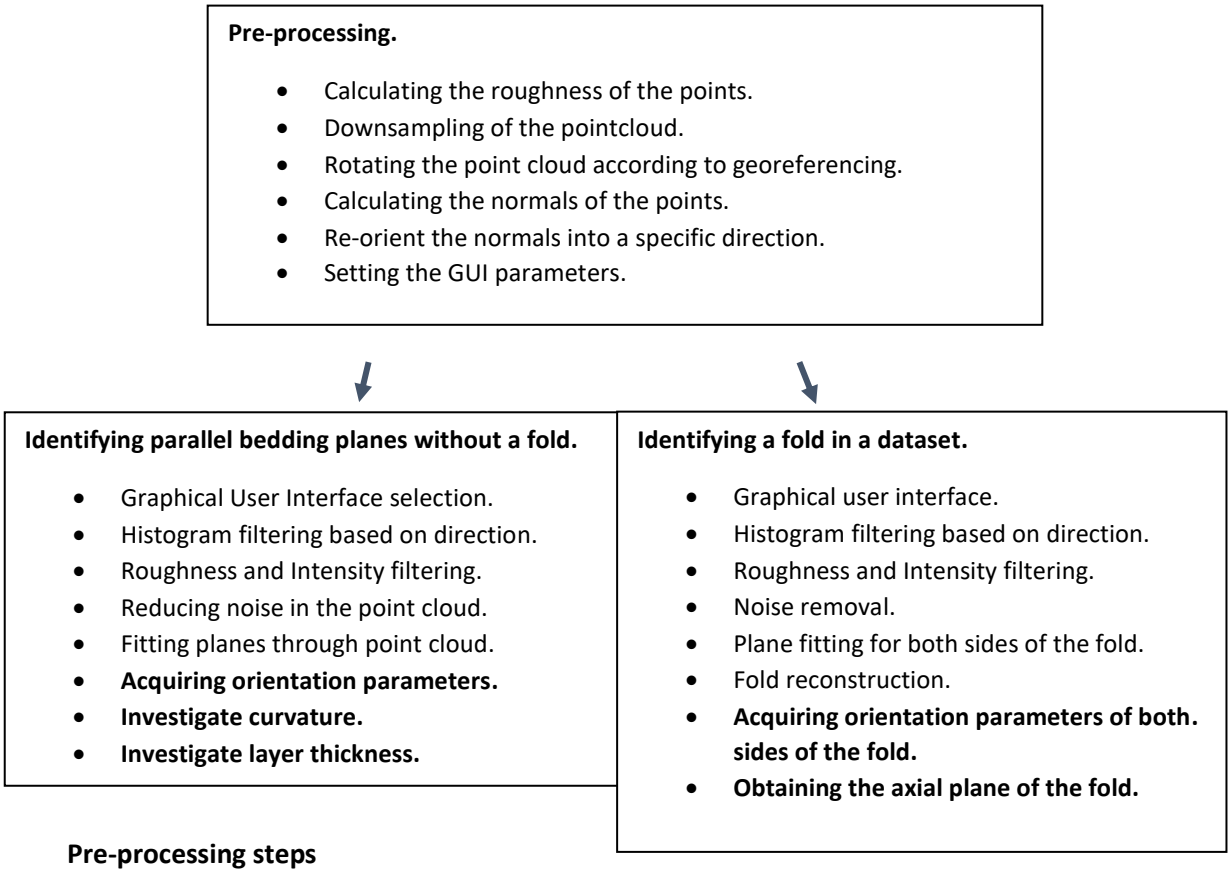
- **How does a successful method look like?**
- **How to efficiently extract the required information from large datasets?**

The second research question extends itself to several chapters, the process flow describes the process and is at the same time the guideline for the most efficient way of acquiring the required information. The specific way the information is acquired is described in chapter 3.2, and therefore it is concluded that this question over the course of several chapters.

This chapter will describe the process flow which leads to the successful method of extracting the desired geometric parameters of a geological outcrop mentioned in chapter 2.1.3. A distinction between pre-processing and actual processing is being made, with at every step a small explanation of what happens. There are bold lines in between which indicate the specific result. First the processing steps are illustrated in a diagram, then a small description is provided. Chapter 3.2 will then give a comprehensive explanation for each step.

There are three bold lines indicating first the pre-processing, then the processing steps performed in case of a normal geological bedding without fold and then the processing steps required if a fold is present.

Within the three subsections, the end result is indicated in bold. These end results are also the results stated in chapter 2.1.3 where the geological features are also mentioned.



Pre-processing steps

1. *Roughness calculation.*

   For every point, a roughness index is calculated.

2. *Subsampling.*

   Data needs to be subsampled in order to work with smaller chunks of data later during the processing.

3. *Point Cloud rotation.*

   The orientation of the point cloud is often not correctly shown when it's acquired. With the use of real GPS coordinates, the point cloud is being rotated and orientated such that it represents reality.

4. *Normal calculations.*

   Normal calculations are needed to determine orientation parameters (dip angle and dip direction) of the geological layers. For every point, the normal is calculated with respect to its neighbors. With normals, computations in the following steps can be performed.

5. *Normal orientation.*

   The normals need to be correctly orientated since at the initial calculation the normals can be pointed into two possible directions.

6. *GUI parameter settings*

General parameters for the graphical user interface are being determined. A subsampled point cloud with a preset size is prepared which can then be called at any moment in time.

In case of a bedding identification without a fold.

7. *GUI.*

   The graphical user interface is called in order to select a number of points which are interesting for computation.

8. *Histogram selection.*

   Histogram based filtering is performed in roughness and (if present) intensity. An estimated direction must be given for the computation to be performed correctly.

9. *Noise removal.*

   Noise removal based on distance parameters is performed in order to get rid of unwanted and non-interesting data. Because histogram and roughness/intensity based filtering has already been performed, distance based noise removal will work very well since only little noise will be left.

10. *Plane fitting.*

    Plane fitting by means of the MSAC algorithm is being performed in order to obtain the geological layers. These plane fitting results are statistically analyzed to **estimate dip directions** and **dip angles**. The results are also analyzed in order to determine if there is a small **curvature** present in the layers

11. *Distance between planes.*

    The distance between the planes is calculated in order to perform statistical analysis on it to determine **layers thickness**. Wrongly fitted planes can be identified and distances between planes can be determined for analysis of the Milankovitch Cycles.

In case of fold identification

7. *GUI*

   The graphical user interface is called in order to let the user define point on both sides of the fold, these points will be used for data filtering.

8. *Histogram selection*

   For both sides of the fold, histogram selection and roughness/intensity filtering is used in order to obtain only the relevant and most likely normals in the dataset.

9. *Noise removal.*

   Noise removal is again performed on both datasets belonging to either sides of the fold. The goal is to remove as much noise from the dataset as possible.

10. *Plane fitting.*

Plane fitting by means of the MSAC algorithm is performed. For the fold condition, much stricter parameters are chosen for fitting planes. The planes need to be much larger in order to be accepted, this because “shopping” can be more of a problem in this situation. “Shopping” is a situation in which points not belonging to an actual plane are grouped and classified as a plane. When “shopping” is happening in a dataset with a fold, it is less likely that the fold will be correctly visualized.

11. *GUI.*

    The graphical user interface is called once again in order to obtain 2 points, one on either sides of the fold.

12. *Filtering and plane fitting*

    Steps 8-10 from the normal plane situation are performed. This includes point cloud filtering based on roughness, intensity and distance. Also includes a histogram selection phase and the plane fitting algorithm for both sides of the fold. These plane fitting results are analyzed to **obtain orientation parameters** of both sides of the fold

13. *Fold reconstruction.*

    The fold present in the outcrop is reconstructed by a series of calculations involving plane intersection calculations and vectorizations. The results include an example of a **potential Fold axis** and **axial plane**.

3.2. Feature estimation

The geological features which are bold in the previous chapter, are obtained by performing certain processing steps. This chapter gives details on each such step, coming to the desired results mentioned in chapter 3.1. These results are extensively elaborated on in chapter 3.2.9 and 3.2.10.

3.2.1. Subsampling

Subsampling is the action of reducing the amount of points within the point cloud. Sometimes this is done out of necessity, and sometimes it is done to reduce the chance on biased results. A bias in the results may appear because parts of the outcrop are more heavily scanned than other parts of the outcrop. The result is that more heavily scanned parts (which are not necessarily the interesting parts) may dominate the results. In this dataset, a bias might appear because the face-on of the outcrop (the cut-off of the layers on the surface of the outcrop) are most likely more heavily scanned than the actual interesting bedding of the layers. This aspect will be discussed and visualized in chapter 4.2. Pre-processing.

There are several types of subsampling which can be performed. The first type of subsampling is random. The program used (which is Cloudcompare) will then pick the amount of points you have defined, at random. This is a chanced based picking and thus will not necessarily solve the bias problem. The Octree subsampling protocol will subdivide the point cloud into an octree, only the point closest to the octree center is kept. Since the maximum octree level is 21, it is assumed that a lot of detail will be lost.
The third and used subsampling method for this thesis is subsampling based on distance. For this method a specific minimum distance is chosen. This minimum distance is the distance between two

points in the point cloud. The protocol starts with a seed point and searches for points which have a minimum distance of x meters from each other (x defined by the user), the found points are also at least x meters from each other, the method continues until there are no more points or no more seed points. This protocol results in a very regular grid like figure 10, dealing with any bias caused by oversampling.

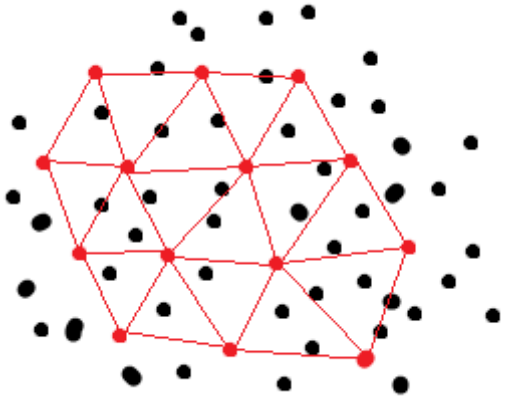


Figure 10: Visualization subsampling protocol. Red lines represent the distance between the points which is a minimum of x meters.

The settings chosen in this thesis is 1 centimeter (0.01 meter), resulting in a dataset in which the distance the distance from any point to any other point is at least 1 cm.

3.2.2. Roughness calculation

For the roughness calculation, Cloudcompare is also used. For this calculation, the only input parameter for every point is the kernel size. The principle of this method is also visualized in figure 11. The kernel size is a sphere within which the protocol searches for points. Through all the points within the sphere, a least squares plane is fitted. The roughness for the center point of this kernel is then equal to the absolute distance “d” between the point of interest and the least squares plane.

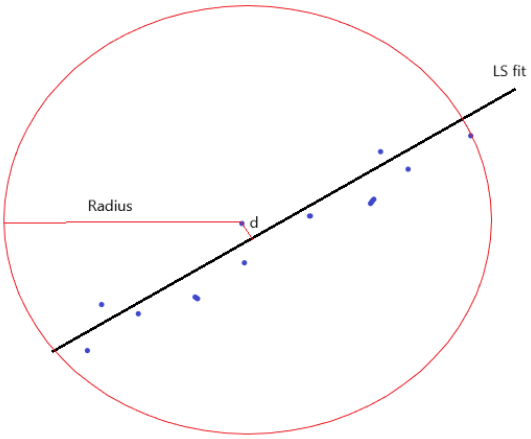
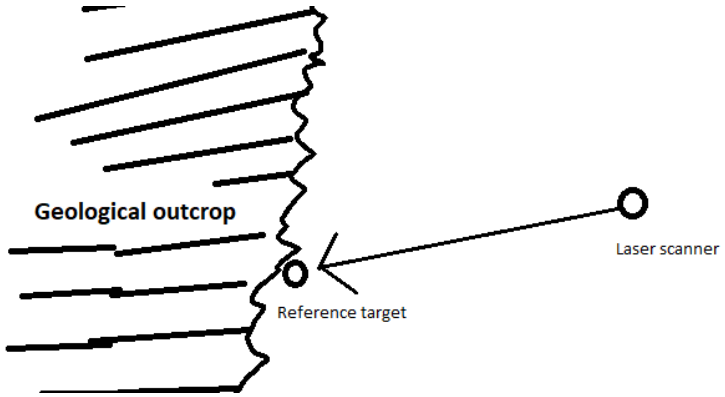


Figure 11: Sketch of roughness calculation.

3.2.3. Point Cloud rotation

The point cloud as it is acquired by the laser scanner, uses local coordinates and orientations relative to the laser scanner. This means that without proper GNSS corrections, the orientation of the point cloud could be in any random direction (the internal orientation of the laser scanner). In order to correct for the orientation, the following method is applied:

A vector is defined from the laser scanner position to a specific target, in this case a target placed near the outcrop such as figure 12a. The position of both the laser scanner and the specific target is defined in both the local coordinates and the real world coordinates (GNSS). The vector defined with the local coordinates is not representative for the real direction but the vector obtained by GNSS is. Therefore the local vector is corrected with the vector obtained by GNSS by the angle difference between the two shown in figure 12b. This difference in the angle is the measure for how much the point cloud needs to be rotated in order to represent the real orientation.



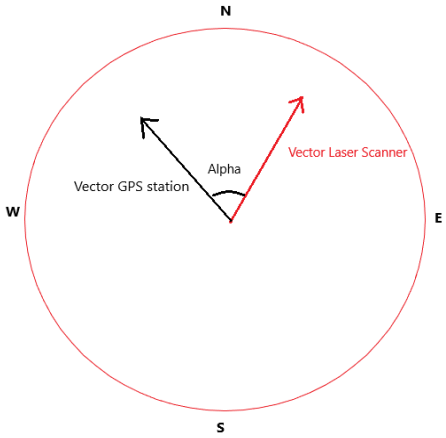


Figure 12a: Reference target illustrated in the geological outcrop.
Figure 12b: Point Cloud rotation. The red vector is the hypothetical vector created with the laser scanner and the black vector is greated by means of the GPS receiver. The angle alpha is the angle over which the point cloud needs to be rotated.

3.2.4. Normal calculation

A very important property of a point is the normal of a point. Without any normal definition, the normal of a point could have any direction. This is however not the case, and a proper normal definition is important for any further calculations. The algorithm used for this is already incorporated in Matlab. The method calculates for every point (by default) the 6 nearest neighbors and fits a least squares plane through those 6 points (illustrated by figure 13a). According to the plane parameters belonging to that one fitted plane, a vector for the normal of that plane is defined. The normal can point in any of the two directions (a normal of a plane can have two direction) and need to be uniformly orientated. For this uniform orientation we choose an upwards direction vectored by the vector [0 0 1]. We orientate all the points in the point cloud towards that direction (shown by figure 13b), if the angle between the normal vector and the vertical is larger than 90 degrees as with angle b, the normal is turned around 180 degrees such that the angle is smaller than 90 degrees. Angle a will remain untouched since it is smaller than 90 degrees. The result is that for normals belonging to planes, the normals are all orientated into the same direction.

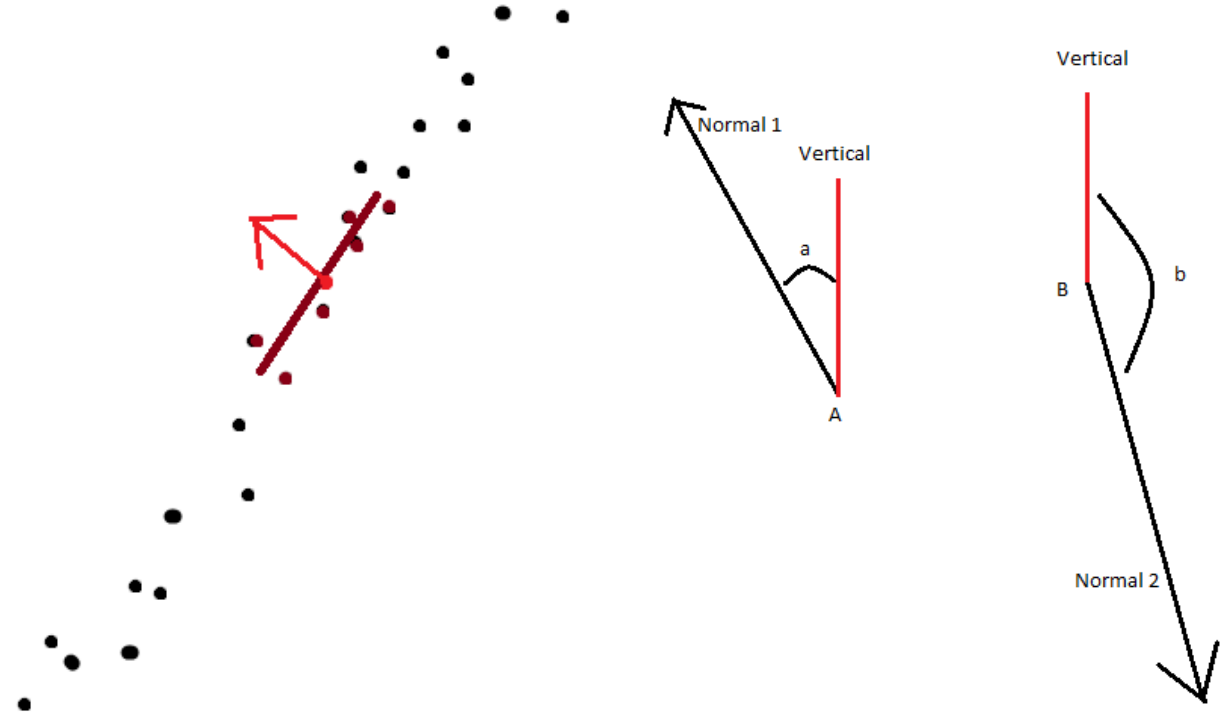


Figure 13a: visualization of the normal calculation protocol.
Figure 13b: The normal of point A is correctly oriented, the normal of point B will need to be recalculated because angle b is larger than 90 degrees.

3.2.5. Graphical user interface

For this thesis, a graphical user interface (GUI) has been implemented into Matlab. Matlab is normally a scripting program, meaning that a preset number and type of calculations and operations are being executed. This offers no room for manual configuration or manual selection of data. This thesis includes a graphical user interface in order to force the user to provide input sampling data used for computations further on. A graphical user interface normally includes buttons, this is not possible in Matlab and therefore a different kind of GUI is made. The graphical user interface includes a plot where the user can select and deselect 3D data points (3D plot) belonging to a point cloud. These selected points are then stored and again reopened at steps further in the script. The selected points will appear in red and the remainder of the points will have the color according to their roughness value (which will also be shown later onwards). Green indicates high roughness values and blue indicates low roughness values. The bedding of the layers are the area of interest for the script, and therefore the user is requested to only select the bedding of the layers. If the user makes a mistake, a point can be deselected as well. The graphical user interface will be shown in chapter 4.2.4.

The visualized 200,000 points are not the points which are going to be used for further calculations. The GUI is only used to select areas of interest as coordinates. Around these coordinates, virtual boxes are drawn within points are going to be stored from the original point cloud which has a size of approximately 14.5 million points. All the points within those virtual boxes will be used in the histogram selection phase in the next chapter.

The function of selecting a point relies on the location the user is clicking. The function captures the position the user clicked, and together with the current angle of the camera, the function decides which point was closest to the mouse click on the screen. This gives possible miss-identification if multiple layers of points are visualized in the same 2D image. Fortunately the low point cloud density reduces that chance to nearly 0 percent. Secondly when a point is chosen wrong, it will most likely be corrected by the sheer size of the statistical analysis.

3.2.6. Histogram Selection

The user input is used to generate a histogram containing information about the selected points (and the points in their vicinity). The normals of the points which have been selected, are decomposed into a dip directions and dip angles. The dip direction distribution of the selected points is made in order to determine which strike direction is dominant in the specifically selected points. This dominant dip direction is going to be a measure for selecting all the potentially interesting points in the whole point cloud. The histogram is going to look like figure 14, where a function based on the top 3 bins will select the normals of interest. The function will not choose the points in the top 3 bins, it will choose at least the points in the top 3 bins. Depending on the position of the second and third best bin with respect to the top bin, a function will determine which points are to be included based on a tolerance. The dominant dip direction will be extracted and the average normal belonging to all those points in the dominant dip direction, will be used for plane fitting later onwards.

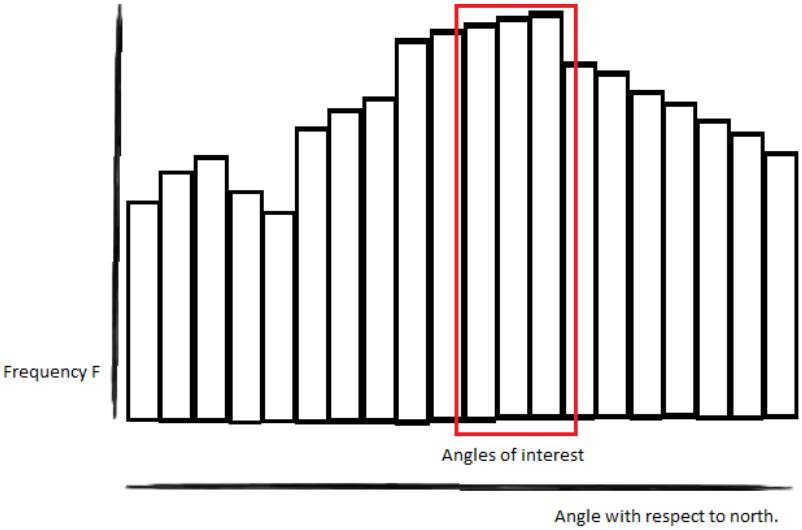


Figure 14: Cartoon style example of Histogram selection

3.2.7. Filtering

There are also two other important properties coupled to every individual point: Intensity and roughness. These two properties are used to further narrow down the potentially interesting data. Points with a roughness above a certain number or with an intensity below a certain number represent data which is of no value to the computation and will only slow the computation down.

Intensity

Intensity values are obtained by a laser scanner (or can be deduced from color values resulting in synthetic intensity values). The intensity represents the reflectivity of the object or surface. The higher the intensity, the more signal is returned to the laser scanner. Low intensity signals often represent very

dark colored surfaces or surfaces which have many different corners and edges. Low intensity observations are less trustworthy and/or also less interesting data points. Therefore, low intensity points are excluded from any further computation.

Roughness

Roughness values are computed by Cloudcompare (as stated in part 3.2.2). High roughness values represent points which are far away from the best fit plane through the data around the point. It is therefore deduced that high roughness values represent corners in the outcrops (as later can be seen in the results) or highly varying local 3D data. High roughness points can therefore also be classified as points which are non-consistent with its surroundings. In figure 15, high roughness points are indicated in green and low roughness points are indicated in blue.

In further calculations the focus will be on consistent data, representing flat surfaces. Therefore high roughness points are of no interest and need to be excluded from further calculations and thus will be filtered out.

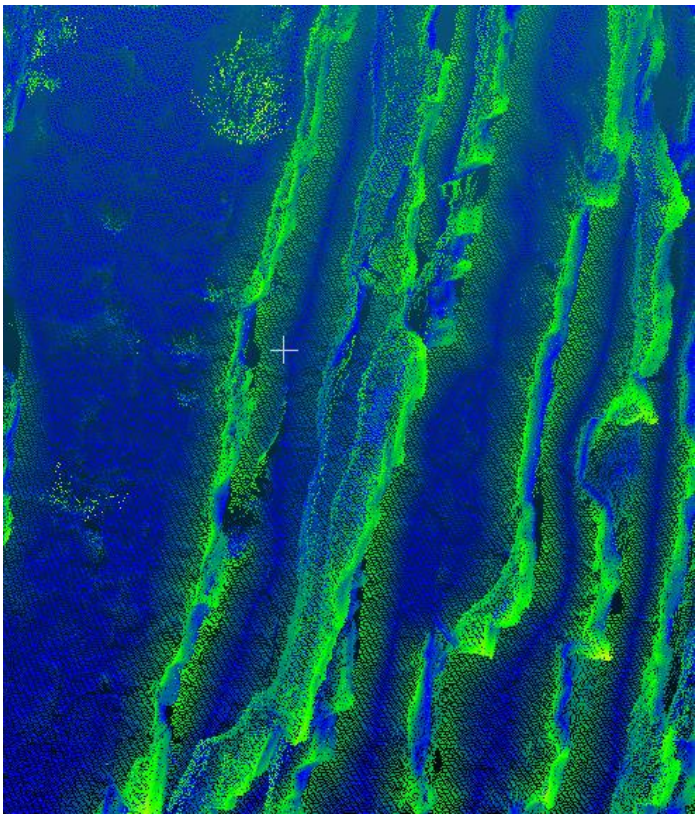


Figure 15: Green values represent high roughness points and blue values represent low roughness points. The points are located at resp. the face on of the layers and the bedding of the layers.

3.2.8. Noise removal

To remove any other noise or unwanted data, a denoise function from Matlab is being used. Any points which are isolated or are still not conform the data format desired, are filtered out in this last step. In this last step, the mean distance to neighboring points and the standard deviation of the mean distance to the neighboring points is being used for filtering purposes.

If the average distance from point of interest “P” is larger than C (C is a scalar) times the standard

deviation of the mean distance to the X neighboring points (X is defined by the user). Then the point is excluded from the dataset. The last noise removal is used to remove small clusters of points and keep larger clusters of points, this because potentially interesting clusters of points are more likely to be large than small. The configuration of this last step will also be focused on preserving only large constant clusters of points.

3.2.9. Plane fitting

After all carefully tuned filtering steps which applicable for most datasets, planes are fitted through the remaining data. For this we use the already implemented algorithm within Matlab. This algorithm is called MSAC (M-estimator Sample Consensus) and is a more robust version of RANSAC. RANSAC (RANdom SAmple Consensus is an iterative method for determining the parameters for a mathematical model belonging to a set of points. An average RANSAC algorithm looks as followed as provided by Fisher (2002):
1: Selecting a random subset of points of the original dataset. These are the hypothetical inliers
2: A model is fitted to this hypothetical set of inliers (this is often a Least Squares model).
3: Test all remaining data on the determined model and determine which points fit the model and which points do not fit.
4: Add the correctly fitted points to the consensus set of that specific model.
5: If all data has been tested. New parameters of the model can be estimated based on the new information. In this way the model can be (slightly) improved.

RANSAC uses a 1-0 switch for any point which is assigned resp. not assigned to fit to the model. Inliers to the model are all considered as equal, resulting in every points adding in the same way to the accuracy of the model. Outliers are also all considered as equal, if the points are further away than the set threshold (set by the user), they will all receive the same constant penalty.

MSAC provides a major improvement on the idea above as stated by Torr (1996), MSAC keeps working with the set threshold. A distance further than the threshold means exclusion and a points closer than the threshold means being included. However the penalty for both cases is no longer constant (1-0 situation), the penalty is now based on the squared distance of the point to the model. The result is that points closer to the model fit get higher priority than points within the threshold but a little further away. The computation time advantage is found in the fact that points beyond the threshold which are still close to the threshold, are more easily considered to belong to the model. Those are more easily considered to belong to the model then points very far away from the threshold. This offers room for improvement because of selective processing of data and thus that is also the reason MSAC is slightly faster than RANSAC.

3.2.9.1 Results for parallel bedding planes.

For datasets without a fold present in the data, this is the last major step required before the eventual desired results can be generated. When no fold is present in the data, the **Layer orientation parameters**, **curvature in the layers** and the **layer thickness** are the desired results. This paragraph describes how these results are obtained after the plane fitting step.

The first desired result is the **layer orientation parameters**. These orientation parameters are obtained by statistically analyzing the planes fitted in chapter 3.2.9. These planes are considered to be sufficient according to the script settings, which leaves only the average dip angle and direction and their

respective standard deviations to be calculated. This is done by decomposing the plane normal into two components: one angle with respect to north and one angle with respect to the horizontal. All planes are included in the statistical analysis, which might include some undesired planes. These undesired planes have unfortunately not been filtered out.

The second desired result is a potential **curvature** which could be present in the geological outcrop. The curvature is not defined as a difference in angle, merely as a representation of the distance of the points with respect to their fitted planes. The colors will indicate what the position of the point is with respect to the plane, and will thus gradually shift between two colors based on their position below or above the fitted plane.

The last desired feature is the **layer thickness**. Layer thickness is defined as the distance between two consecutive planes. This distance is defined by the perpendicular distance between the bedding of two following geological layers. This distance can be visualized with figure 16.

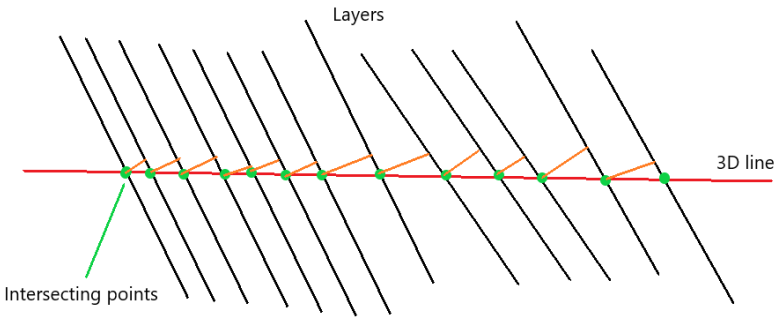


Figure 16: distance between layers.

This distance will be evaluated along a 3 dimensional line (visualized in figure 16). This 3 dimensional line crosses all fitted planes at approximately the center of the fitted planes. The center of the planes is chosen since it is believed that the center has the smallest potential error. The intersection between this line and all the planes will be calculated and these intersections are chosen as the base to calculate the distance to the neighboring plane as shown above.

The quality of the results is dependent on the quality of the fitting algorithm. Ideally one would like to acquire all bedding planes present in the outcrop, unfortunately this will not be possible. An estimate of the expected amount of planes can always be given, being a goal to acquire. The further the result is from the actual amount, the less reliable the result is.

3.2.9.2. Results for datasets with fold.

For datasets with a fold, a combination of the previously used functions is used with some new functions. Combining the graphical user interface together with the filtering steps results in the **orientation parameters of the layers on either sides of the fold.** The dip direction and dip angle will be

known for both sides of the fold after plane fitting. Plane fitting will be performed according to the GUI data obtained by the user input. The user will have to define areas which the script has to search in for both sides of the fold. Thus, the user is automatically defining both sides of the fold, resulting in the orientation parameters.

3.2.10. Fold Reconstruction

Fold reconstruction is an action which consists out of many small operations and pieces of data.
One important part to mention is that this fold reconstruction is only going to work when the user selects the same part of the same geological layer, on both sides of the fold. The figure below illustrates what is being mentioned. The user needs to select either the bottom or the top of a geological layer at either sides of the fold. If the top of the geological layer is selected at one side and the bottom on the other side, the fold reconstruction will fail and thus the fold reconstruction is reliant on the user. As long as the user picks a point on the same side of the geological layer, the fold reconstruction will be accurate. The principle of point selection for this stage is visualized in figure 17.

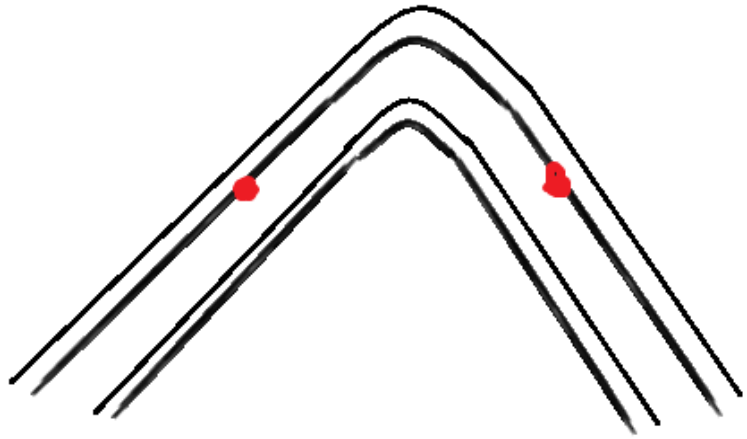


Figure 17: Picking points at the same side of the geological layer, at opposite sides of the fold.

For both sides of the fold, most of the geological layers or at least the most important geological layers has been identified. For both sides the most dominant average normal is determined. A plane is now fitted through the two manually chosen points with their corresponding dominant normal. This results in two planes which represent both sides of the geological layer. In order to determine the fold plane (which incorporates all the information) a point on the fold plane and the two vectors spanning the 3D space of the fold plane, need to be determined.
The point on the fold plane and the first vector are determined by intersecting the two planes belonging to the two sides of the fold. The result is a line with a direction in the third dimension, this is also called a fold axis. The direction of fold axis is the first vector which spans the fold plane. The point on the fold plane can be determined by solving the equation for one specific value of X, Y or Z.
The second vector needed to determine the space of the fold plane is extracted by calculating the bisector of the two normals used for the plane definition for both sides of the fold. The bisector between the two normals is the “average” vector of the two, splitting the space between the two. The fold plane or **axial surface** will be located exactly in between the two normals, and thus this second vector is the last vector required to span the space of the **axial plane.**
To summarize, the bisector of the two normals belonging to the planes on both sides of the fold,

together with the second vector and point created by the intersection of those two planes, span the space of the **axial plane** present in the data.

3.3. Software and hardware

Processing is implemented on a windows machine using Matlab. A script made by Elizabeth Prentice (2017) will be used and improved. The script is made more efficient such that it is capable of processing more data in the same time. Then the script is stripped to only maintain the parts which are potentially interesting for this thesis. Potential software such as Cloudcompare and Meshlab have been considered to produce intermediate results.

Of all possible commercial, software, Cloudcompare and Matlab are used for this thesis. The machine on which the script is run has the following specifications:

- OS: Windows 10
- CPU: AMD FX-8350 Octacore at 4.0 Ghz
- GPU: NVIDIA GTX 1070
- RAM: 16 GB at 1866MHz.

Hardware configurations are mentioned since they affect the behavior and speed at which computations and their results are obtained and visualized. The CPU determines the speed at which processing is taking place. The GPU determines how well and how fast the results are visualized on the system and the RAM determines the total capacity of the computer to temporarily store data and the speed at which the data can be exchanged and accessed.

3.4. Data requirements

In order to make the point cloud data useful for computation, specific requirements need to be meet. The level of which these requirements are met determine the quality of the results, therefore the below mentioned points answer the following question:

- **What parameters influence the detectability of the geological features and to what extend?**

1. *As little artefacts as possible.*

   An artefact is an unwanted piece of information. Artefacts can come from the laser scanner in the form of sunlight. Sunlight can sometimes be seen as a laser reflection, resulting in a line of points pointing towards the sun. Other artefacts are bushes, trees, grass or other rubble incorporated in the data. All these artefacts need to be dealt with since they do not belong to the data itself and will therefore influence the processing by making it slower (more data to process).

2. *Include rock bedding surfaces.*

   In order to identify rock layers, the surface of the rock layers need to be scanned. Without proper scans of (ideally both sides of) the geological layers, no correct orientation parameters can be identified. Measurements taken parallel to the strike line are visually appealing, but not very useful with respect to the goal.

3. *Visible folds.*

Next to visibility of the surfaces of the rock layers, visibility of the features which are supposed to be recognized is also required. The fold needs to be visible in the point cloud. They need to be visibly present in the geometry of the layers in order to be recognized by this computational method.

4. *Measurement geometry*

   The measurements cannot be taken too far away from the outcrop. Measurements too far away result in a very low point density. Also measurements taken under large angles (parallel to the detection surface), result in low point densities and bad measurements.

All the above mentioned requirements have their own influence on the detectability of the features. If an outcrop is totally covered by vegetation, one will unlikely be able to extract any valuable geological information. The other above mentioned points also have their significant impact on the detectability of any geological parameters. If there is too little data on the dataset, not enough data will be available. And lastly if there is no actual depth showing the bedding of the layers, no results can be obtained since this method relies on the 3-dimensional geometry of the outcrop.

Summary

Acquiring the desired result cannot be done if the data is not correctly being taken care of. A lot of computational tricks are incorporated in this method in order to diminish the chance on errors and to reduce computation time as much as is reasonable and possible. Some actions such as subsampling, roughness calculations and normal calculations have to be done on the bulk data. The actions performed after the Graphical User Interface are more refined and are aimed to perform calculations on a very specific set of data, filtering the dataset of undesired data. These filtering steps and the GUI are the computational tricks applied to this dataset in order to make the method work. Apart from the software and its implementation, the machine hardware influences the speed at which the results are obtained. Faster processors with more cores at higher speeds and more physical memory at higher speeds will certainly decrease computation speed.

Apart from the computational tricks, the dataset characteristics is one of the most important considerations before the whole computation starts. If the data requirements are not met, the computation cannot be performed.

4 Results

When all the methods and features have been described and the methods used to obtain them are mentioned, the results should be evaluated and visualized. This chapter includes a data description for the two outcrops in order to show what datasets have been worked with. The second part of this chapter will include the results obtained.

The proposed processing flow in chapter 3.1 together with all the feature calculation steps in chapter 3.2 result in all the features which have been identified. Chapter 4.2-4.4 will describe the results obtained from the feature extraction with in chapter 4.2 the results of the pre-processing and chapter 4.3-4.4 containing the results belonging to bedding plane extraction and fold extraction.

4.1. Data description

For this thesis, a specific set of scanned outcrops is used in order to evaluate the outcrop, design a method and to design the method such that it can be used for multiple outcrops (if certain conditions are meet). Below, a description will be given concerning the datasets used.

4.1.1. La Charce

The “La Charce” dataset is the most exemplary dataset currently available. The dataset is acquired by Elizabeth Prentice during the 2016 Bsc. fieldwork in France. It is located in the French Alps near the small village of La Charce (figure 18a). It is an outcrop of a geological rock formed in the Cretaceous. Within the Cretaceous, the rock was formed in the Hauterivian era. The layers are approximately 135 million years old. One can see from figure 18b that there are two types of layers. These two types of layers are limestones and marls. Limestones consist of mainly calcium carbonates and is relatively strong compared to the marl layers. The marl layers consist out of clays which are considerably weaker and less resistant to corrosion.

The clear distinction between the strengths of the two layers is visible in the two images of figure 18b. Here one can see that there is a clear distinction between the stronger layers on the foreground and the weaker layers on the background. The weaker Marl layers tend to corrode faster and therefore the stronger limestone layers are the layers of interest. Their bedding planes remain intact and can be analyzed when observed by a laser scanner.

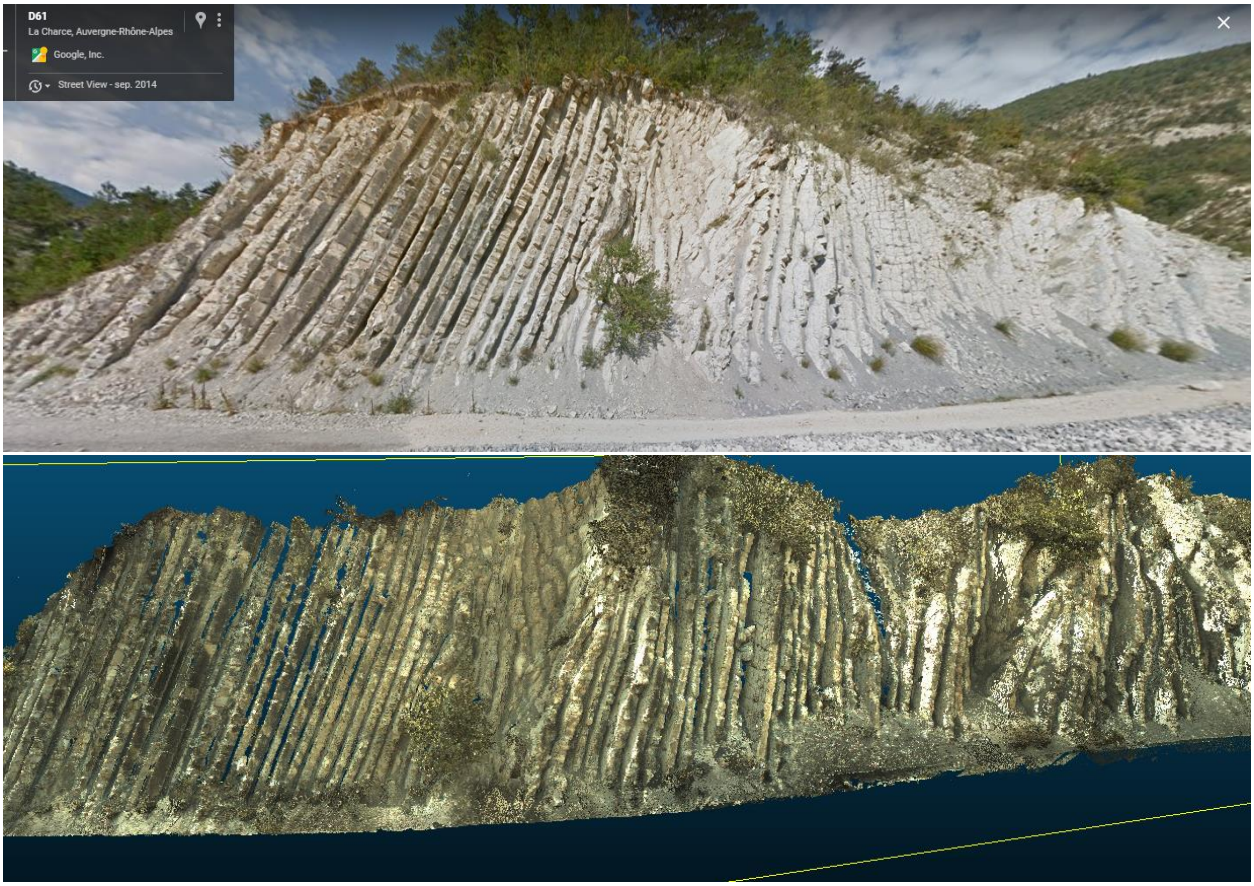


Figure 18a: Location of the La Charce outcrop. Source: Google maps
Figure 18b: La Charce outcrop via google maps and point cloud based.

There are three main types of data which are acquired: GNSSP (Global Navigation Satellite System Positioning), Photos and Point Clouds. For 2 out of 4 locations a laser scan is made with color photos, for the 2 scans, also GPS measurements were acquired.

The previously acquired dataset was acquired by Elizabeth Prentice in 2016 in the BSc. fieldwork in France. Data was acquired with two different systems: A Leica C10 Scanstation and a Trimble 5700. The Leica C10 Scanstation was used for making point clouds and photos and the Trimble 5700 was used for gathering Global Navigation Satellite System (GNSS) data which provides information about the location.

The Leica C10 Scanstation's main purpose is obtaining 3D point clouds of the environment. The wavelength of the laser of the laser scanner is a 532nm wavelength green colored laser scanning up to 50,000 points per second. The camera is used to color every points resulting in a very detailed sample of the environment containing x,y,z values together with RGB information and even intensity information of the returned laser signal. The internal camera has a 1920x1920 pixel resolution (4 megapixels) with a 17x17 degrees incidence angle, these RGB values are used to color the point cloud

The used GNSS receiver was a Trimble 5700. It was used on every outcrop at 1 measurement in order to provide information on the location and orientation in the coordinate reference systems. It is a 24-channel duel frequency receiver capable of receiving signals from GPS, GLONASS, Galileo, Beidou and

other regional systems. Acquiring positioning of 2 scans made is crucial for orientation calculation purposes.

The in figure 18b illustrated dataset has the following properties:

- Original 69.2 million points reduced to 17727727 points.
- Point density is about 1 point per cm^3
- A total of 4 laser scans have been made
- The scans includes X, Y and Z location information. RGB color information and Intensity information
- Approximate size of the outcrop: 100 x 20 meters

4.1.2. Meso-scale chevron fold.

This (outcrop) dataset will be called the “Meso-scaled fold” and is located in the northern Apennines in Central Italy acquired by Amerigo Gorradetti et al. It is a formation originating from the Mesozoic era [1]. Its specific age is unknown, it’s estimated to be anywhere between 70 and 250 million years old. This outcrop contains layers of rock which are mainly limestone, they are reasonably well preserved. It is unknown if the outcrop is georeferenced. However, computations and calculations can be performed even if the point cloud is not geo-referenced. The only problem is now that the results cannot be verified,validated or compared.

The precise location of the dataset is unknown, the approximate location is given in figure 19:

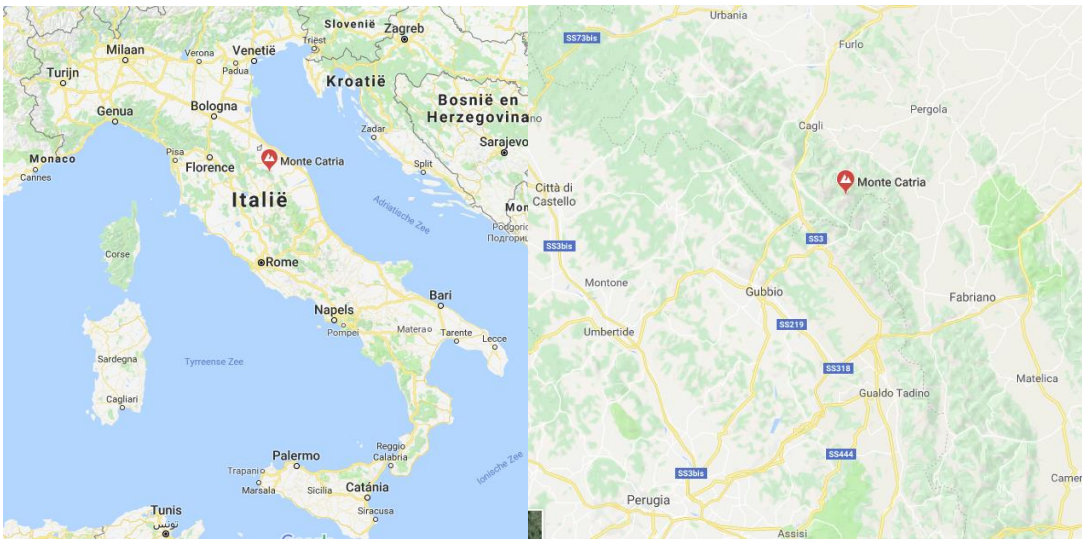


Figure 19: Apennines.

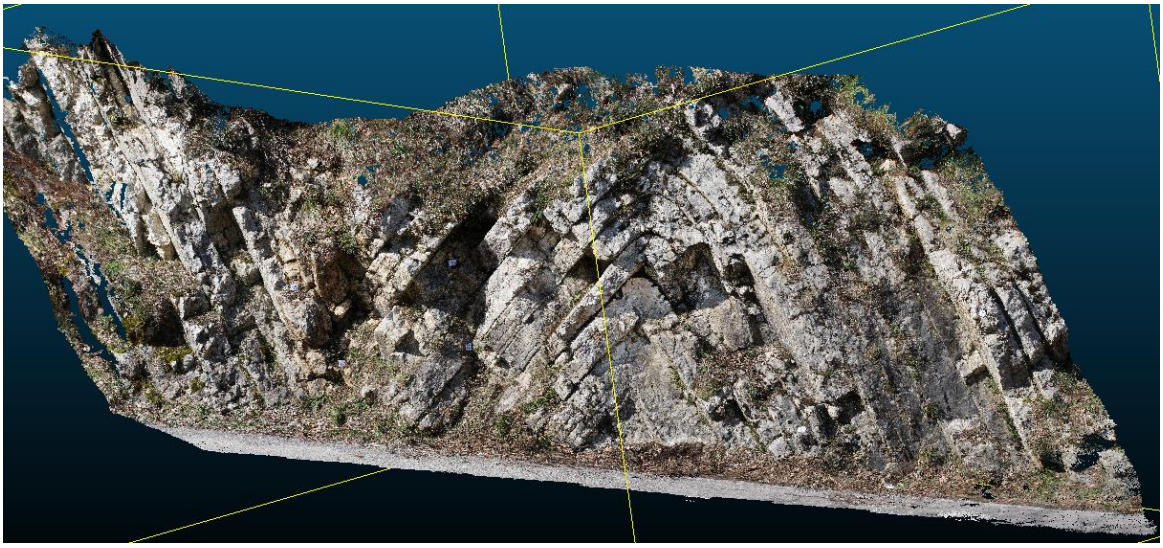


Figure 20: North Apennines double fold.

The above illustrated dataset has the following properties:

- Original point cloud size is 23,023,087 points
- Point density is about 50 point per cm^3
- A total of 351 photos have been made
- The scans includes X, Y and Z location information and RGB color information
- Approximate size of the outcorp: 12 x 4 meters

This dataset and all the information is acquired from Corradetti (2017).

4.2. Pre-processing results

The subsampling is meant to reduce the amount of points in the point cloud. The reason for reducing the amount of points is not to reduce calculation time but to make the point cloud more uniform. When the point cloud is more uniform, the chance of acquiring biased data is smaller. Biased data occurs when there's a difference in data sampling across the outcrop. Figure 21 shows the original and subsampled point cloud. Figure 21a shows an oversampled face of the outcrop and figure 21b is subsampled.

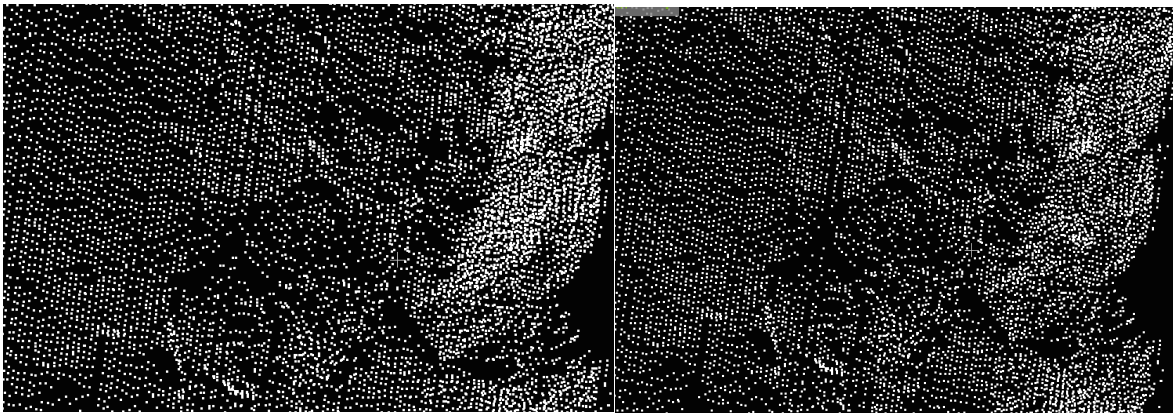


Figure 21 A: original point density point cloud.
Figure 21 B: Subsampled point cloud.

In Figure 21 A, the point cloud in its original density is visualized. The white points on the right side of the image are the face-on of the outcrop and the darker points to the left is the bedding of the layer.

The whole point cloud is being subsampled, but this subsampling affects areas in a different way. The face-on of the layers is often seen by several scans (often 2 or more), while the bedding is mostly only seen once. This is the reason why the face-on of the layers is often affected the most by subsampling. In figure 21, the two situations are visualized. The left image shows the original point cloud and the right image the subsampled point cloud. It is not very clearly visible, but there are less white points on the right side of figure 21b than there are on figure 21a. Also the left side of the images is largely unaffected by the subsampling.

The subsampling was performed with Cloudcompare, and the chosen subsampling method is the equal distance subsampling method with a minimum spacing of 0.01 meters. This means that the minimum distance between two arbitrary points in the point cloud is at least 0.01 meter (1 centimeter). The subsampling results in a reduction of about 2.8 million points from 17.3 million points to 14.5 million points or a reduction of approximately 16 percent. Most of this point reduction follows from a reduced density at the face-on of the outcrop.

The roughness calculations are important for identifying less interesting parts of the outcrop (at least for certain calculations). The roughness increases for areas where corners and edges are present in the outcrop. This is also visible in figure 22.

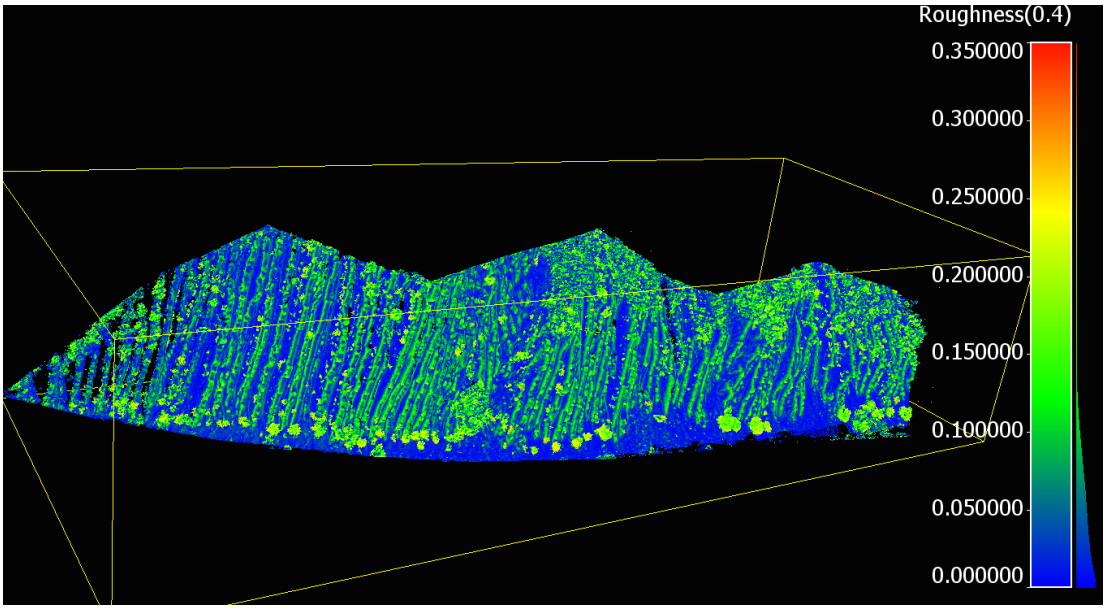


Figure 22: Roughness calculation with a kernel size of 0.4 meters.

There are very distinct color differences visible in the outcrop. The lighter the color the higher the roughness values. Green and yellow colors indicate roughness values from 8 centimeters up to 36 centimeter. This means that the maximum distance from the center point and the fitted plane for the 0.4 meter kernel around that point is approximately 36 centimeters.

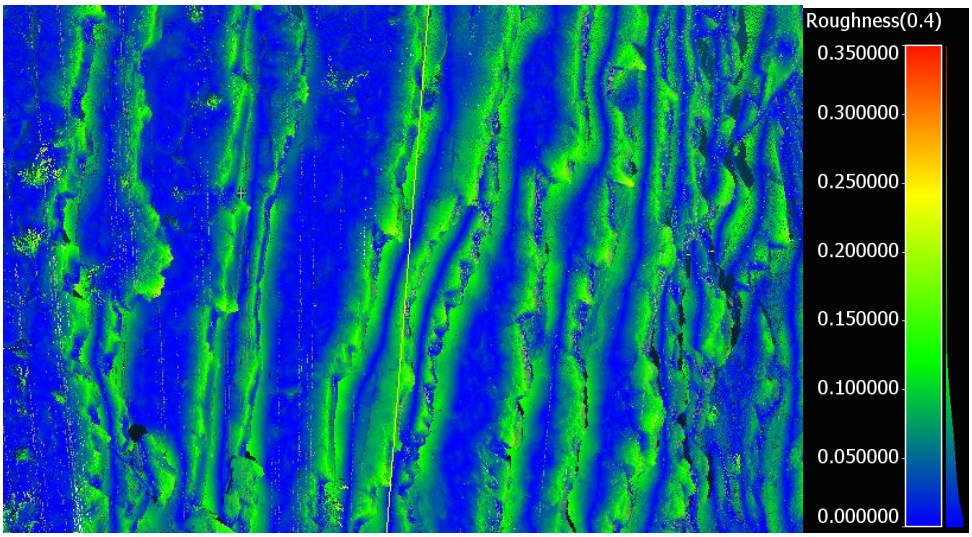


Figure 23: example of commission error for roughness calculations.

The advantage of this large kernel sizes is that corners and rough areas are very easily filtered out since the roughness is high. But also areas which are supposedly correct are filtered out and identified as too rough. An example of this is shown in figure 23.

In the image shows the face of the layers is very clearly visible in the green areas, and are therefore also very easily filtered out for any further calculations. Unfortunately, one can see that there is a color transition from bright green to bright blue, and this color transition is introduced at the edge of the face-on. This results in good and potentially interesting areas possibly being filtered out. This is however not an feature which can be solved at the roughness filtering step since it may result in the face-on not being filtered out as well. Even though potentially interesting data might be filtered out, enough data will be left.

The normal calculations done with Matlab is the most important source of data for all calculations. These normals are used in most selection steps including selecting dip direction for layers and estimating parameters for identifying a possible fold. Chapter 3.2 described how the calculations are done. However, it also describes that the orientation of the normals can have two possible directions (one is exactly the opposite of the other). The two images in Appendix I show resp. the original and re-oriented point clouds. The normals have been decimated in order to better visualize the results in a plot. A sample of this whole outcrop has been taken in order to better visualize the reorientation of the normals, these two samples are shown in figure 24.

One can see clearly from figure 24A and 24B that for the indicated plane, a more consistent orientation of the normals is visualized. It has to be noted that the same normals are visualized in both images, only the orientation of the normals in the first image is not consistent and therefore a part of the normals disappear behind the point cloud. In the second image all normals are pointed in the same direction and therefore more normals are visible.

Figure 24A: original orientation of normals
Figure 24B: recalculated orientation of normals

It is important to have all the normals pointing into the same direction. It is necessary because it creates consistency in the data. The consistency results in the planes being more easily detectable in later steps. Later on, a histogram is used in order to determine the dominant dip direction of a selection of points. Having dip directions which vary too much and have no consistency result in a histogram based selection which will not be representative. Figure 25 is a histogram as it selected by the Graphical User Interface in step 4.2.4. The histogram selection is shown in this chapter makes a selection of points based on what the user thinks is the dominant dip direction. The custom made program will then select the dominant dip direction based on a search radius of the selected points and a histogram as is shown in figure 25 is generated. The dominant dip direction is determined by making a histogram with a reasonable resolution, the histogram resolution is not too high with a bin width of approximately 6.5 degrees it would represent greater accuracy than actually possible if a histogram would be used with a bin width of a few degrees. The histogram for the La Charce outcrop including all filters and all selection stages looks like figure 25 where a part of this histogram will be selected for further plane fitting.

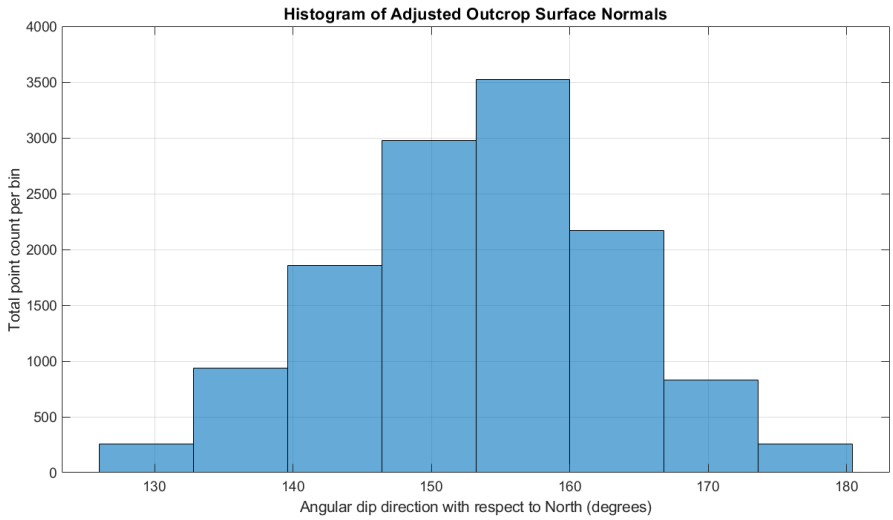


Figure 25: Histogram selection based on GUI input.

The graphical user interface for the La Charce point cloud (which has no folds) will look like figure 26. It has no distinctive buttons as is normally expected from a GUI. The graphical user interface is a plot with every visualized point being interactive and clicking a point in the plot will make the point turn red. The user can rotate the point cloud and zoom as he/she likes in order to better visualize and identify the preferred layers.

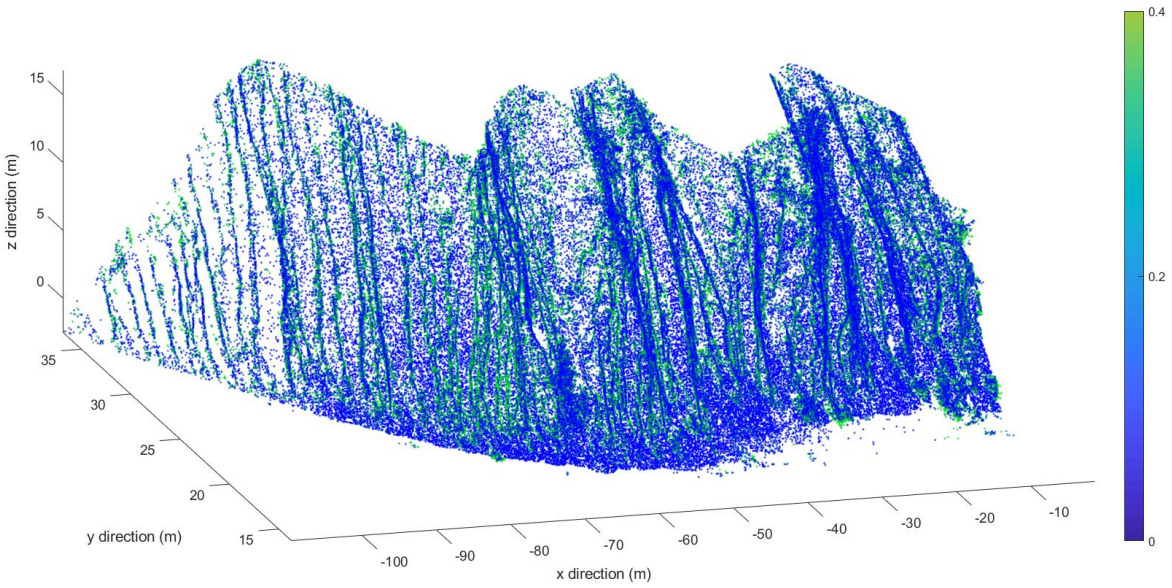


Figure 26: The Graphical user interface.

A small subset of points from figure 26 has been made in order to illustrate what it looks like when points have been selected.

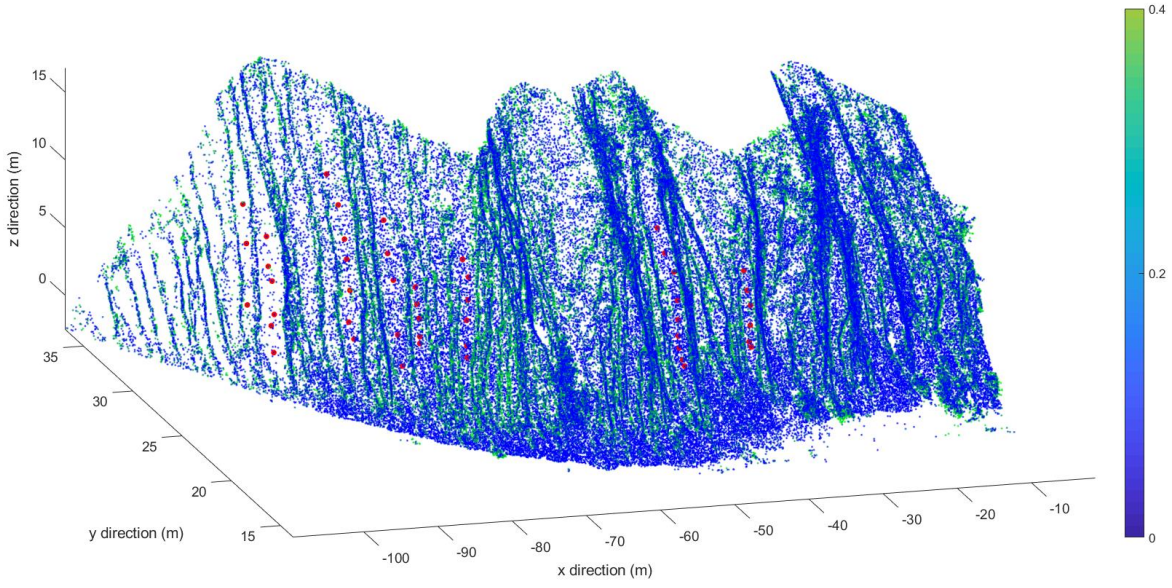


Figure 27: Selection of points, an example of what the GUI looks like.

In figure 26 and 27, the same colors are used as in chapter 4.2. The colors in this plot are the same as the roughness values. The point density in this plot is significantly lower since the functions used for

plotting these points are not optimized for point clouds in Matlab. Unfortunately there was no other way of representing the points, and thus the amount of points visualized is approximately 1 percent of the total amount of points present in the point cloud (200,000/14,500,000). The figure contains a custom color bar indicating a range from 0 to 0.4. It is a custom color bar since the default color bar option within Matlab does not provide a good representation of the values since the color bar option looks at the color values of the plot and not the intended roughness values (which are combinations of colors).

4.3. Extraction bedding planes

Noise removal is a step which is performed after the manual and histogram based selection. When the points are manually chosen and only the points with a desired dip direction are left, there are still a lot of points which do not naturally belong in the data. E.g. a local clusters of data which coincidentally have the same direction because of a block of rock in the outcrop. In order to get rid of loose, noise removal is performed with Matlab. The "pcdenoise" function does this in this case. The method has been described in the previous chapter, the parameters chosen are not close to standard values. Since large clusters of points are important, and continuity is important, a relatively strict filtering procedure is taken which filters out small individual clusters very well. The chosen "X" nearest neighbors to point "P" is equal to 50 with a threshold C of 0.024 meters. The combination of these two parameters works very well for very dense point clouds with a minimum spacing of 1 centimeter since the desired layers are mostly preserved while noise is mostly removed. These parameters have to prove themselves in other point clouds. This evaluation will be performed later on. With these parameters, the filtered point cloud has been obtained shown in figure 28. One can see the global image obtained after filtering. The red points are the points which have been filtered out based on the filtering properties. The green points stay naturally because the cluster density they're in is high enough to keep them. The high point cluster density is typical for planar areas.

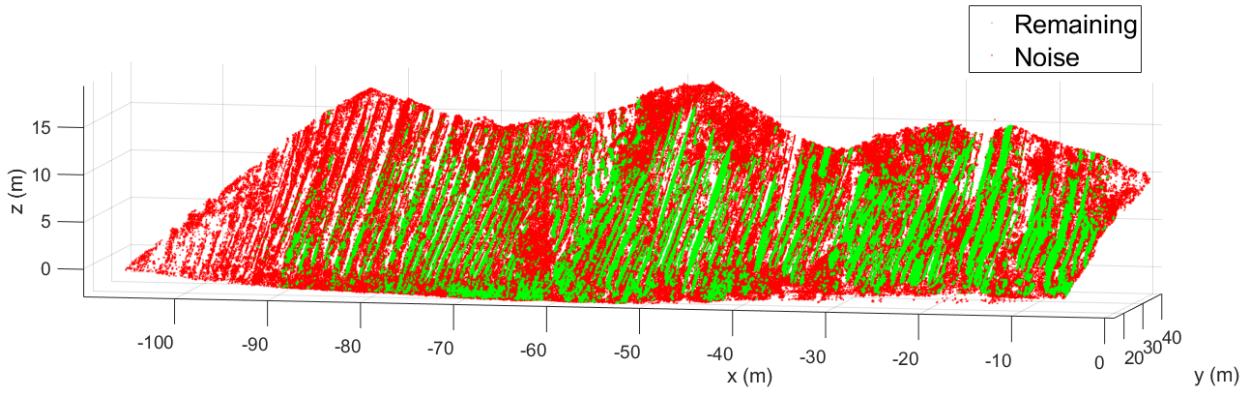


Figure 28: Noise removal global image. Red are points which are removed because they are classified as noise. Green points are kept since they are not classified as noise.

A close-up of a small area in the point cloud is shown in figure 29. This image shows better what happens with the filtered points. After all the steps performed before filtering, points not belonging to any plane in the data. These points are often just individual points, but they greatly affect the planes fitted to the data later on in the process. Figure 29 shows the close-up with remaining points in green and points which are filtered out in red (just like in figure 28). One can see that the filtering step has a

positive influence on the potential detectability of the planes in the next step. The planes are largely untouched while noise is largely filtered out. Unfortunately this filtering step also filters some planes out of the data. This is visible in the same figure 29, where the clear plane in the middle emphasized by a blue indicator is filtered out because it does not qualify the criteria. More of these planes are filtered out as can also be seen in figure 28 on the left hand side left of the x=-90, there are clearly some planes visible and those have been removed automatically. And these planes are unfortunately filtered out which is not desired. When the filtering parameters are set less strict, these planes will be included in the data, but this will be together with the unwanted noise which has previously been successfully filtered out. The choice has been made to filter out some small layers as well instead of incorporating extra noise, this since extra noise will greatly influence the plane determination and fitting algorithm.

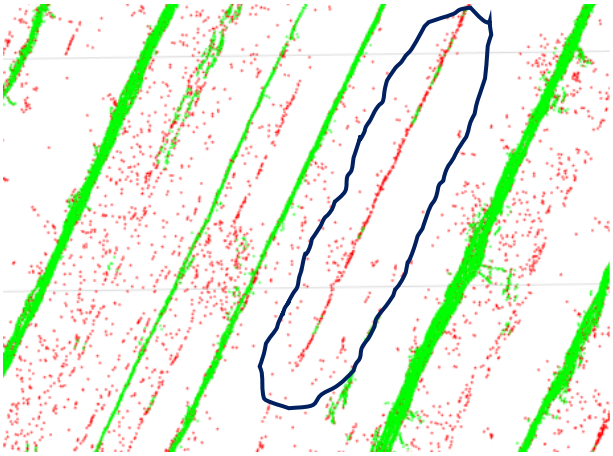


Figure 29: close-up of the filtered layers

On the filtered point cloud, plane fitting can be performed. The plane fitting algorithm produces the following results with respect to the fitted planes. It has to be mentioned that the two images below are not the same result. The same function has been used to create both images, but the normal used for setting the plane definition is different. Figure 30A is made with the normals calculated from the GUI selection phase. Figure 30B has been produced with a different normal. The second normal is derived from the first plane fitting iteration, where an average over all normals has been taken as input for the second iteration. This second iteration has been performed in order to attempt acquiring a better fitting result. The parameters used or the second iteration are the same as for the first iteration, only the

normal used for plane fitting has been changed slightly with a maximum plane fitting tolerance of 1 degree more.

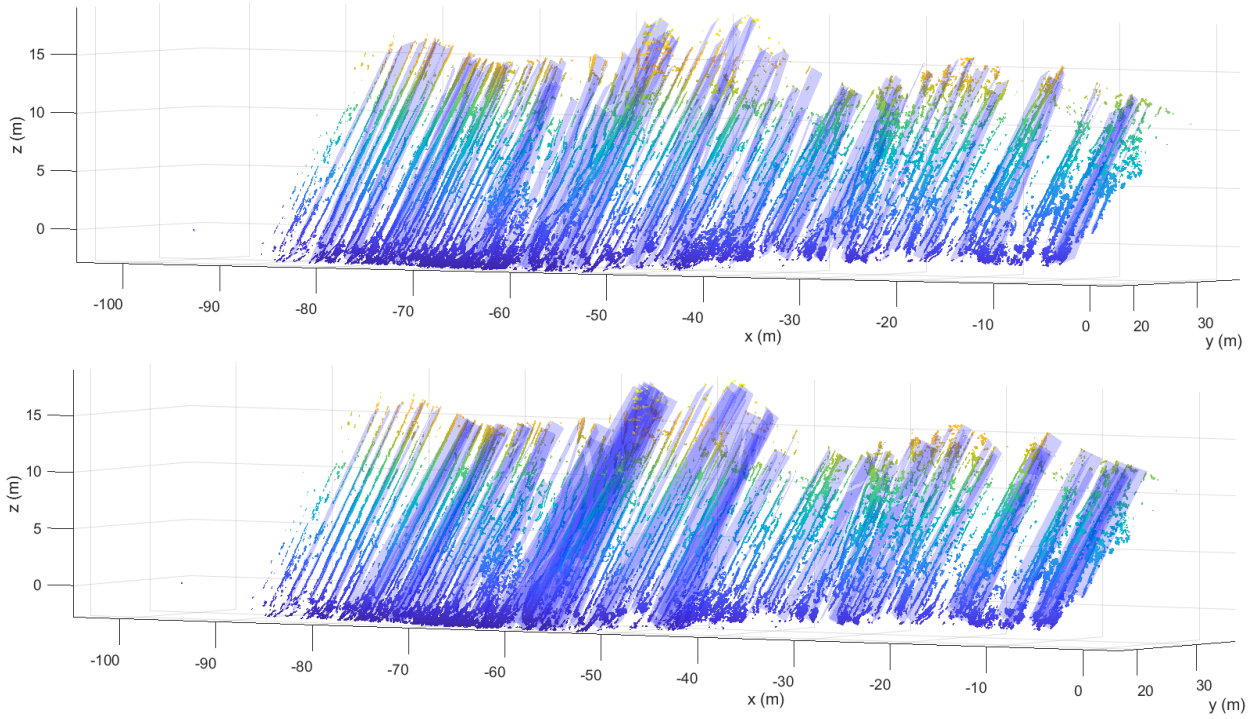


Figure 30A: original plane fitting.
Figure 30B: Second iteration plane fitting.

The best way to visualize the planes is by showing the point clouds belonging to the planes and plot them all in a different color. The axis of the plane have been cut off to improve on the visibility of the point cloud. The viewing angle and the used plane fitting result is the same as in figure 30 A.

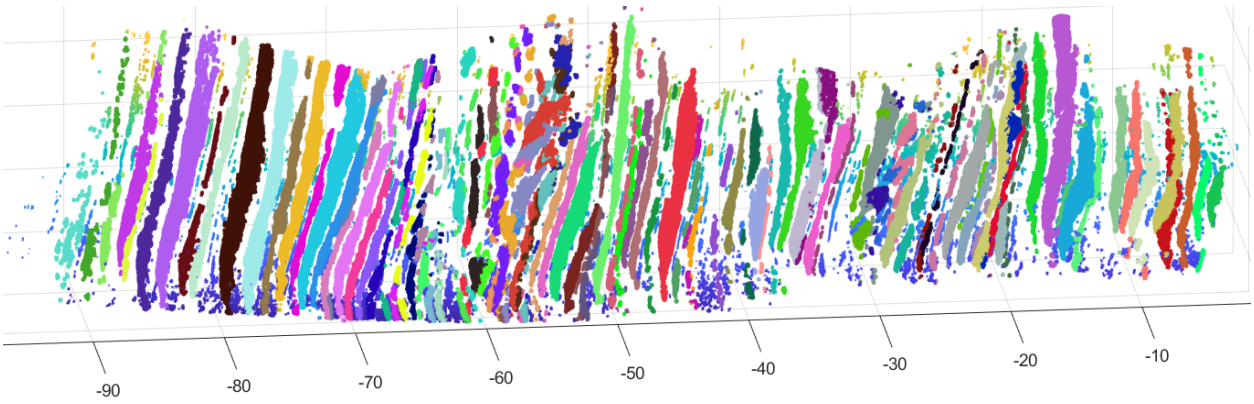


Figure 31: point clouds belonging to the planes. All are given a different color to better visualize the contrast.

One can see from figure 31 that the individual planes which are present in the data, are very well identified by the plane fitting algorithm. Most bedding planes in the actual outcrop are also identified by the plane fitting algorithm. In figure 31, also the original point cloud fed into the plane fitting algorithm is shown. One can see that in the bottom part of the image, a lot of data is excluded together with a lot

of data being falsely included in plane data. The bottom +/- 1.5 meter is mostly rubble which has passed through all filtering steps before. If that data is sufficiently fitting the already present plane, it will just be included in the fitted plane.

The results obtained above are also analyzed with respect to orientation parameters belonging to the planes. Dip direction and angle of the fitted planes will indicate what orientation parameters belong to the real-life geological planes. The following table will summarize the findings regarding the orientation parameters for the first and second plane fitting attempt.

Table 2.1. First plane fitting attempt

Plane fitting 1	Average	Std. deviation.
Dip direction (degrees North)	156.68	2.45
Dip angle (degrees)	52.38	0.76

Table 2.2: Second plane fitting attempt

Plane fitting 2	Average	Std. deviation.
Dip direction (degrees North)	155.96	3.08
Dip angle (degrees)	51.89	0.90

One can see from table 4.2 and 4.3 that the results of the first and second attempt are very similar. The dip direction and dip angle of both fitting attempts are within 1 degree while the standard deviation of the orientation parameters do the same. In the first attempt the algorithm has an angular margin of 5 degrees while the second attempt has an angular margin of 6 degrees. This accommodates for the slight difference in the results (for both the orientation parameters and the errors).

The largest problem in the obtained result is that there are planes falsely fitted to noise data, also so called “shopping” is occurring where points belonging to different geological layers are joined to become one fitted plane. These errors are called commission errors: falsely identifying planes. The largest commission errors are found in figure 31 around -50 to -60 (on the x-axis). This area represents a discontinuity in the outcrop where slightly different plane orientations are present, resulting in misleading and wrongly fitted planes.

Apart from commission errors, also omission errors occur in the results. These omission errors mainly occur in the left hand side of the image, where some small planes are not identified. The amount of commission errors are far greater than the amount of omission errors. Unfortunately these errors cannot be avoided and it is believed that the current state of the program is providing the best tradeoff between the two types of errors.

Of course omission and commission errors can be excluded and filtered, unfortunately they cannot be filtered out at the same time. Decreasing omission errors increases the amount of commission errors and vice versa.

From the planes derived during the plane fitting algorithm, one can derive more information than just the planes themselves. The following image shows the distance between consecutive planes. The planes

have been sorted by drawing a line on the height one wants to calculate the distance. The planes have been sorted along that line.

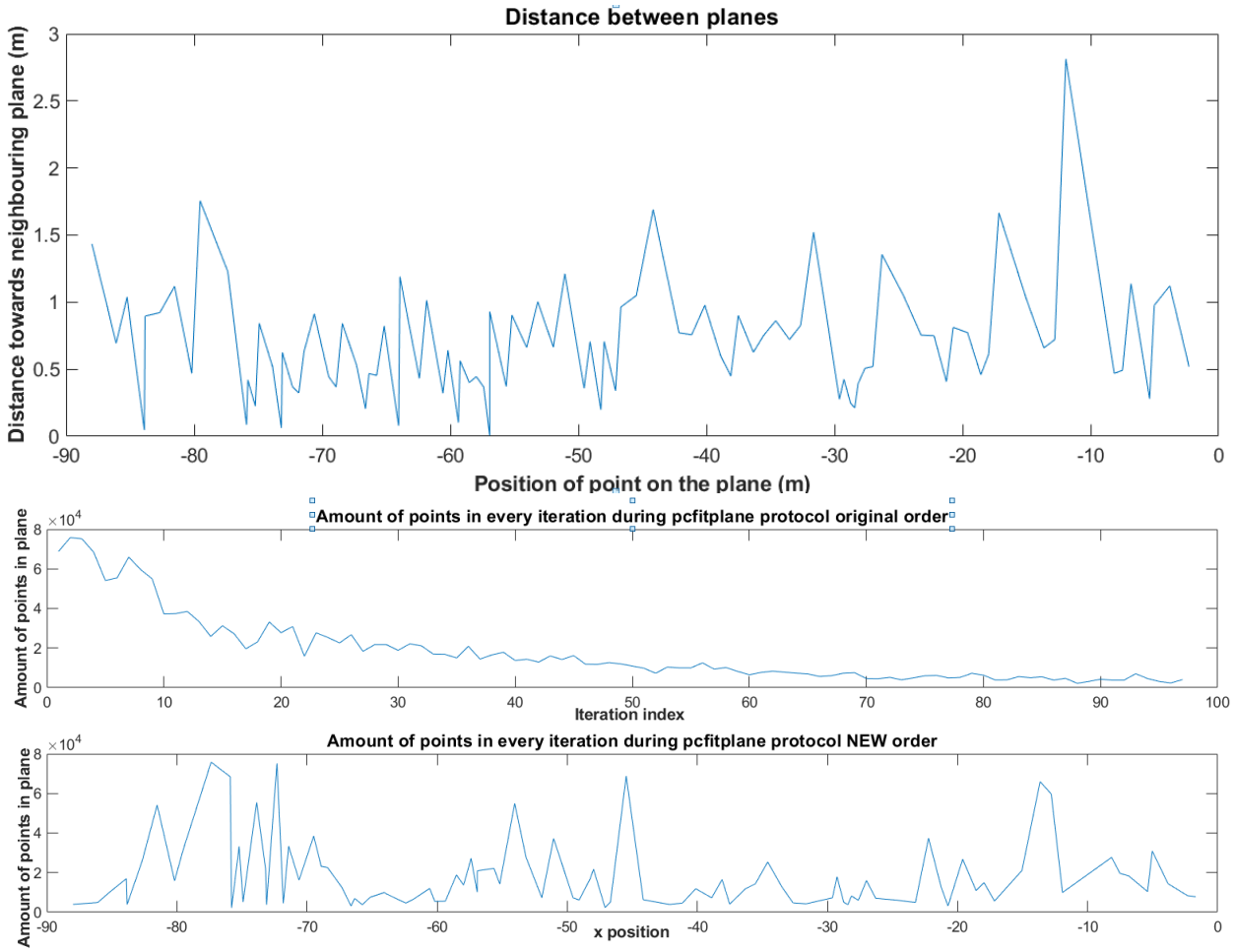


Figure 32a: Distance between consecutive planes
Figure 32b: Amount of points in the plane in the original plane fitting order
Figure 32c: Amount of points in the plane organized on x-position.

Figure 32 shows that the majority of the distances is located between +/- 10 cm to +/- 90 centimeters. There are some exceptions of almost 0 meters and greater than 1 meter distance between the planes. The planes represent the top and bottom of the bedding of the geological layers, where no distinction is made between the top or bottom of the bedding layers. These geological layers alternate in thickness from about 50-60 cm to thinner 20-30 cm layers. Larger distance of +/- 50-60 centimeters represent the thicker and stronger limestone layers while the 20-30 centimeter thinner layers represent the thinner 20-30 centimeter marl layers. The few exceptions where the layers are smaller than 10 centimeter or larger than 1 meter represent gaps in the data or the variability in the fitted planes. Figure 32b and c Show the size of the point clouds. Figure 32 b shows the size of the point cloud in the order they are identified, the biggest point clouds are identified first and the smaller point clouds later. Figure 32c shows the point cloud organized and sorted on their x-direction just like figure 32a. One can see that in the areas which are commission errors are abundant, the point cloud size is in general small. Areas which are well identified contain point clouds with higher point counts.

The histogram in figure 33 (made from figure 32b)supports as well that there is a separation of layers, there are 2 major peaks: around 35-50cm and around 75 centimeters. According to previous work by prentice (2017) and Bisschop (2017), these thicknesses belong to marl and limestone layers. Of course, there are also marl layers thicker than 35 centimeter and limestone layers thinner than 50 centimeters but those exceptions will probably not dominate the results.

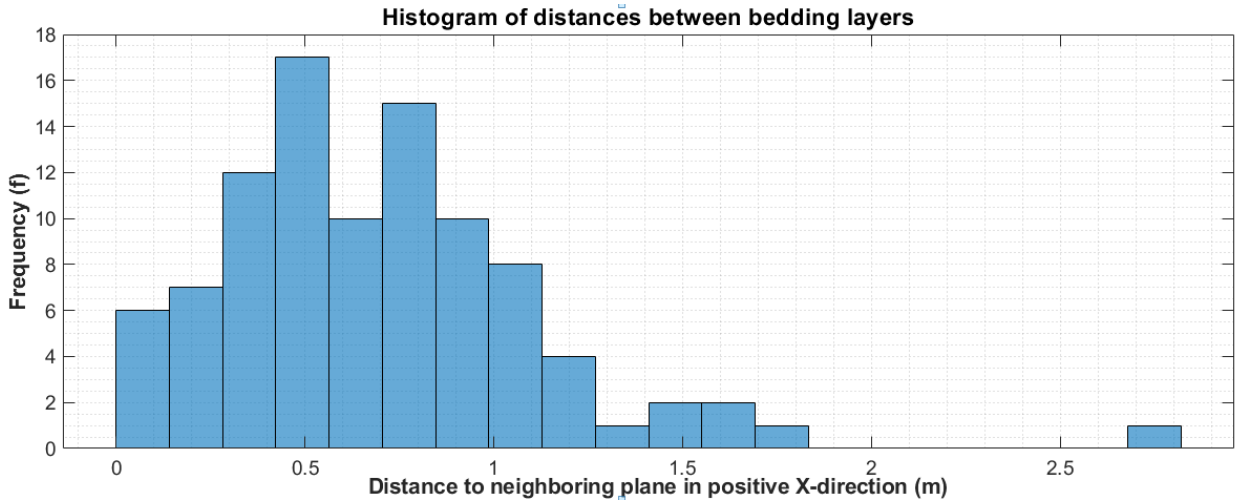


Figure 33: Histogram of distances between the fitted planes

From the original point cloud obtained, it was suggested that there might be a small curvature in the geological layers. The distance from each point belonging to their fitted plane has been calculated and assigned to the point cloud as a color. Red colors in figure 34 below indicate that the point is above the fitted plane (in the z-direction) and green colors indicate the point is below the fitted plane (in the z-direction). The brightest red represents +15 centimeters from the plane and green colors -15 centimeters from the plane.

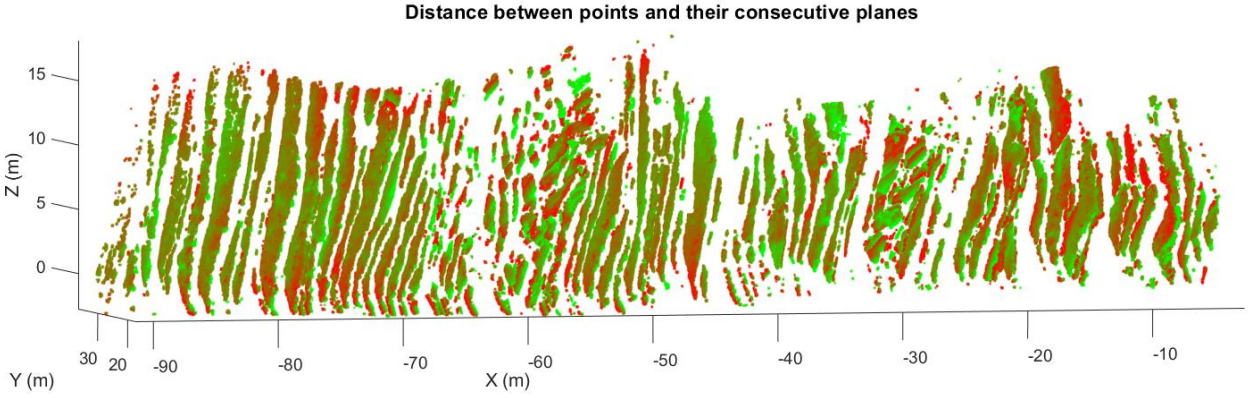


Figure 34: distance from the plane.

Unfortunately, this representation of the distance gives no result with respect to any clear and consistent curvature being present in the results. We do however see a gradual contrast of green and red colors within some fitted plane. This large color contrast is most likely due to the forced plane fitting to the data. Since there is a predefined normal used for fitting planes, any misfit of the used normal can cause for a misfit of the plane and thus a grade in the distance between the model and the actual data.

There are some instances in figure 34, which might prove that there can be a curvature in the rock bedding. Some layers show red colors at the bottom and top, and green colors in the middle. This indicates that the actual bedding of the layers is curved around the plane. 2 layers have been selected which could indicate that this curvature is actually present in figure 35. One can see that in the center 2 planes, the color of the points in the center of the fitted plane is green while at the ends the color is red. This means that in the center of the fitted plane the actual points are below the fitted plane (in the Z-direction) and at the top and bottom of the visible bedding layer the points are above the fitted plane (in the Z-direction).

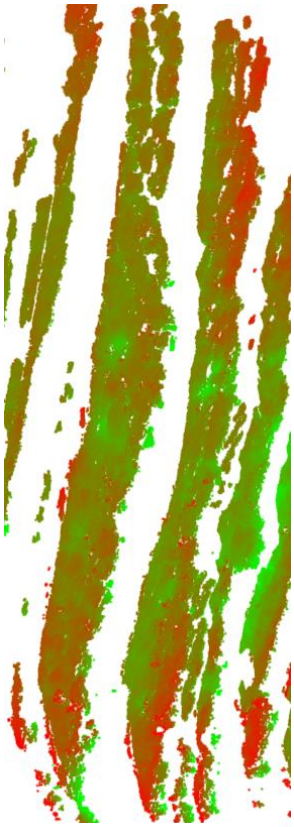


Figure 35: Layers which might imply a curvature

Since not all layers show the same behavior it is unclear if this is actually true, these two layers are only a few layers which could proof its true. According to these two layers, there might be a slight curvature of about 1 degree maximum. This angle is derived from the fact that the maximum distance from the plane on the top and bottom of the layer is approximately 15 centimeter maximum. This results in a deviation of 30 centimeters between the bottom and the top of the plane (a distance of 20 meters). When calculating a the angle the layer would deviate (simple tangent angle), this would be approximately $\alpha = \tan\left(\frac{0.15}{20}\right) = 0.8\ degrees.$

This is of course not the real deviation in the planes angles between the bottom and the top of the bedding layers, the calculated angle is merely an indication. The real angle will then approximately be somewhere between 1 and 2 degrees.

Another interesting aspect of this planar distance plot is that the commission errors with respect to fitted planes are very well visible. Figure 36 shows an area with commission errors, this specific illustration shows one specific and clear case of such an error. In figure 36, one specific identified plane is illustrated in green. The original point cloud is also visualized and one can clearly see that the identified point cloud is a gathering of points coming from several bedding layers. Every visual gap in between the consecutive cluster of green points represents a jump in the bedding planes.

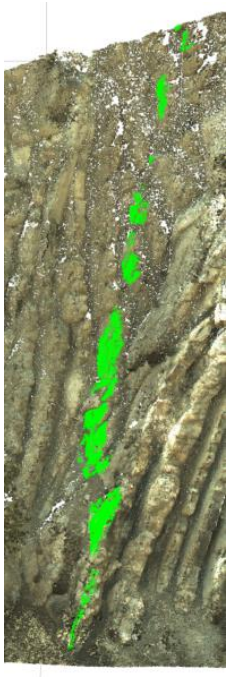


Figure 36: Example of commission errors in plane fitting algorithm.

This distance plot can be a measure to filter out commission errors with respect to fitted planes in future research. This distance plot has unfortunately not been used to evaluate the correctness of the planes during the course of this research.

4.4. Fold reconstruction

The fold reconstruction consists out of several pieces of code which have already been used before together with new constraints and additional code. The graphical user interface (GUI) has been used to make the user selected areas of interest. Together with the histogram selection, a stricter noise policy and a stricter plane fitting algorithm a more strict result has been generated for this specific desired result. The original point cloud is much denser than the previously investigated point cloud. It has a point density which is about 50 times higher than the La Charce dataset, since the point cloud is treated the same, the point cloud will be subsampled with the same 1 cm spacing constraint.
The results regarding the graphical user interface and all the filtering steps is shown in figure 37 together with the result of the points picked during the graphical user interface interaction.

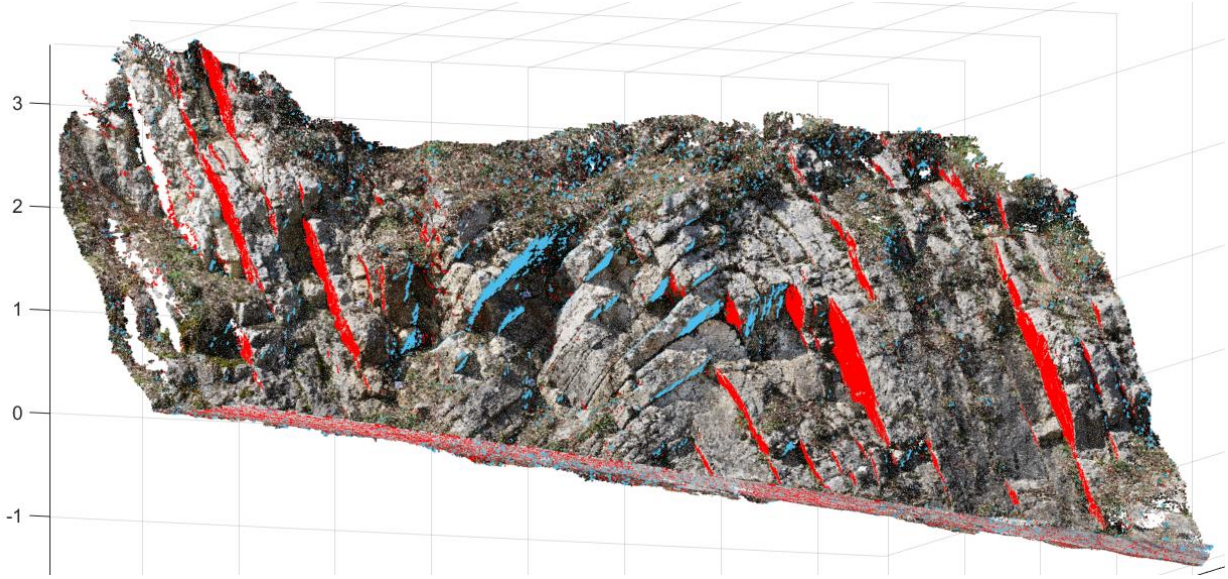


Figure 37: fold limbs identification. Light blue colors indicate the identified first side of the fold bedding layers and the red colors indicate the other orientation of the bedding layers at the other side of the fold.

One can see in figure 37 that there are two preferred point normal orientations present in the denoised dataset. This is according to the general way the program has been set up. The program is written such that only 1 fold can be recognized at a time and therefore takes two defined orientations as input. Unfortunately the script is not able to make a distinction between the layers which belong to both folds present in the dataset. The red points in the left hand side of the image belong to a different fold than the red cluster on the right hand side of the image. It is thus interesting to mention that the orientation of those two red clusters is exactly the same though both cluster belong to a different fold. For the eventual identification of the fold itself this does not matter since the estimated fold location is done by picking two points on both side of the fold.

The plane fitting procedure is the same as usual. However, the minimum plane size requirements have been decreased from 2000 points per plane to 1000 points per plane. This in order to prevent commission errors, omission errors will certainly occur using these settings. However, since identifying all planes is not the main priority, omission errors are not a significant problem. The main goal of this method is to find the fold plane and therefore missing out on some planes is just a minor problem.

After plane fitting, the fold is identified by combining the normals belonging to the dominant dip direction and angle acquired at the graphical user interface step with two points which are certainly located at the same side of the plane bedding. The result from this calculation is the fold imaged by figure 38.

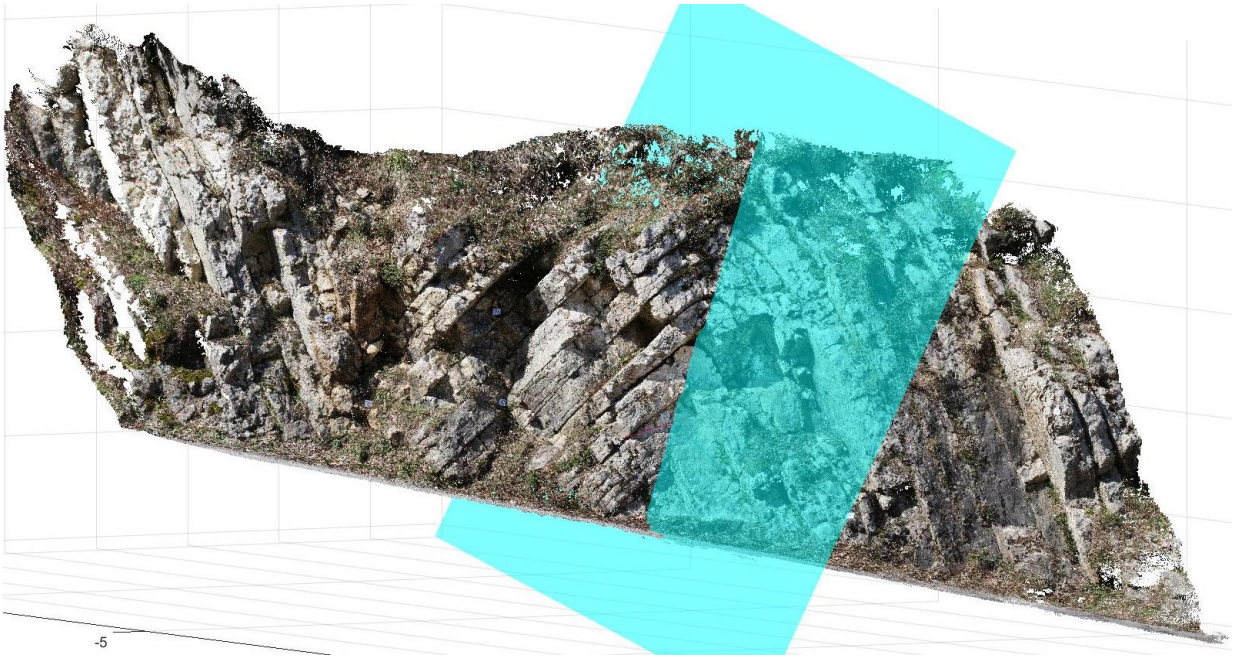


Figure 38: fold plane identification.

Figure 38 shows the fold plane as identified by the method. It is clearly visible that the fold plane is correctly located in the point cloud. The fold is constructed from the identification of the bedding layers on both sides of the fold. Adding two more points identifying the same layer on either sides of the fold is the last piece of data needed to accurately reconstruct the fold according to this method. Figure 38 shows an axial surface which is slightly off, this is caused by a point which has not been correctly chosen. Apart from slight inaccuracies based on the user, the method is still very accurate and gives an axials plane which is almost correct.

Summary

The results are satisfying with respect to how much the script can solve, a lot more interactivity has been added than initially expected. Because of this interactivity, one is able to acquire much more accurate results and is less dependent on automatic detectability. Unfortunately omission and commission errors are still present in the results. These are hard to tweak out of an individual dataset since every dataset is different and thus will respond different to setting changes. The program can be more automated than it is at the moment, and it has to be done at a later moment in time (as also will be mentioned in the recommendations). The added functionality of recognizing a fold is relatively easy to use but does require some minor knowledge of what is going on. The result is shown as a plane, but also more advanced information such as location in space and a formula can be acquired easily from the results.

5 Discussion

The results in chapters 4.2, 4.3 and 4.4 show the results obtained by the process chain of the algorithm. However, these results are not necessarily right. The question rises how these results can be validated and justified. For this matter we can ask the following question:

How to validate the estimated parameters computed by the computer?

In order to answer this question, a comparison is done between the work obtained in this thesis and the work obtained by Prentice (2017) and Bisschop (2017) who have been working on the same "La Charce" dataset. Then a comprehensive list of parameters and settings is given for any function used, and after that a sensitivity analysis is performed on the important settings.

5.1. Comparison to Bisschop (2017)

Timo Bisschop has also been working on the so called "La Charce" outcrop. Timo manually selected 2023 points in the model which have been analyzed by a program called LIME to create the results he obtained. The advantage of the program is that LIME is made for analyzing rock outcrops. The disadvantage of using this method is that choosing the amount of points is very time consuming. Choosing over 2000 points of interest can take possibly 1 to 2 days for one outcrop. This point picking is considered to be processing time which means that it takes at least 1 to 2 days to process this outcrop. In contrast to this "computation time" is the accuracy. The results obtained with this program considered to be fairly accurate because of the incredible data density.

For the dip direction and the dip angle, Timo found that these were respectively 149.01 degrees and 66.08 degrees with a standard deviation of 30.69 and 6.25 degrees respectively. The calculations on the layer thickness shows that there are limestone thickness measurements and marlstone thickness measurements. Figure 39 shows a part of the results of these calculations. The limestone layers are on average a little thicker than the marlstone layers, the histogram in figure 40 supports this. The limestone layers are on average 0.4 meters thick while the marlstone layers seem to be approximately 0.35 meters thick, no real estimation of the average layer thickness has been done By Bisschop (2017), only the distribution if given. The conclusion for the layer thickness is that on average the limestone layers are slightly thicker than the marlstone layers.

These results are very similar to the results obtained in this thesis. The obtained orientation parameters are a dip direction of 156.68 degrees and a dip angle of 52.38 degrees with a resp. standard deviation of 2.45 and 0.76 degrees. The dip direction results are compatible with each other but the dip angle results are differing a lot. The layer thickness calculations are arguably the same when comparing figure 33 with figure 40. Some layers in this thesis are not identified, resulting in in a histogram with less counts and larger gaps. On average the distances are very similar.

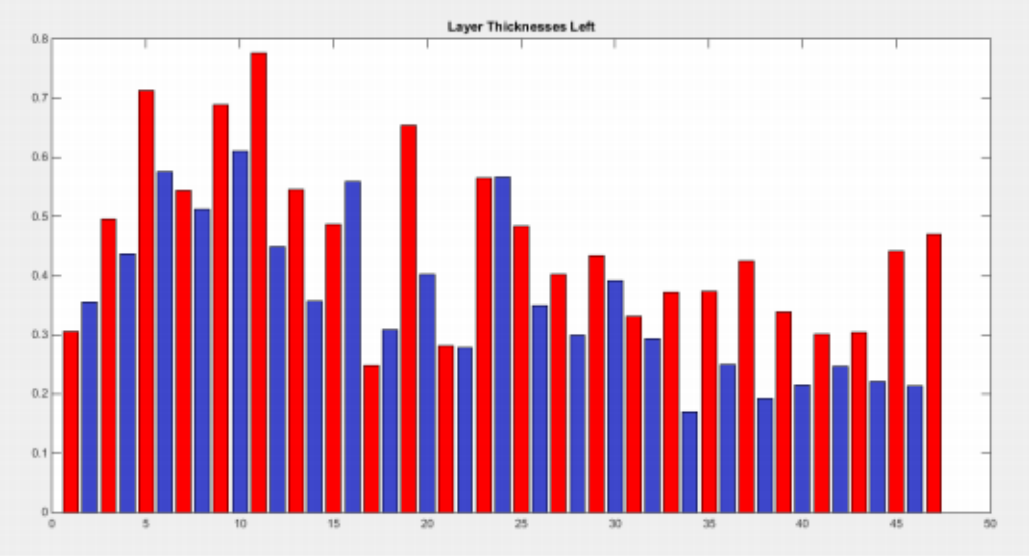


Figure 39: Limestone (red) and marlstone (blue) layer thickness of the one part of the left part of the outcrop

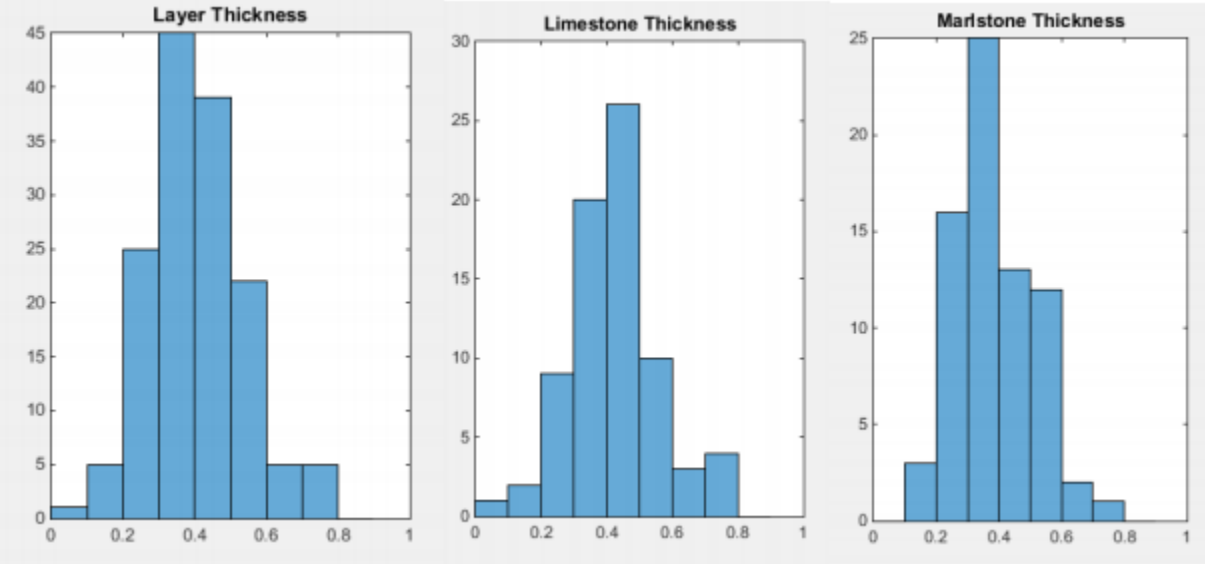


Figure 40: histogram of layer thickness (in meters).

5.2. Comparison to Prentice (2017)

Elizabeth Prentice also worked on the La Charce dataset. She used a code she wrote in Matlab in order to evaluate a part of the dataset. Prentice's result included a part of the outcrop roughly 2 meters by 2 meters of the outcrop. This part consisted out of approximately 300000 points. This is only about 1.6% of the total amount of points present in the outcrop (300000 points evaluated /18700000 points in the whole outcrop). This computation gives some insight on the outcrop, but unfortunately only 16 layers have been evaluated (8 marlstone and 8 limestone) which means that the results lack the numbers. Her estimation of the dip angle is 64 degrees with a standard deviation of 0.6 degrees and her dip direction is 157 degrees with respect to north with a standard deviation of 1.4 degrees. These standard deviations are considerably lower, resulting from the minimal amount of points and the processing of a very stable part of the outcrop.

The layer thickness calculations results show that the layers in her part of the outcrop show a thickness of 32.17 centimeters with a standard deviation of 6.27 cm for the limestone layers while the marlstone layers are 32.68 centimeters thick with a deviation of 3.29 cm. The layer thickness of both limestone and marlstone are very similar as also can be seen in figure 41. The disadvantage of these layer thickness calculation is that they are done with some pre-knowledge implementations. This means that the layer thickness has been estimated in advance and is used to identify the layers. This means that there is a possibility that if individual layers are wrongly identified because of a discontinuity of some sort, any other layer identification might be wrong.

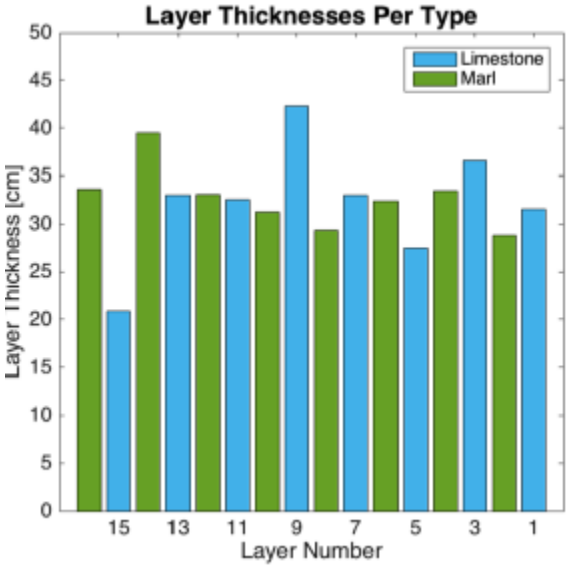


Figure 41: layer thickness for Limestone and Marl layers.

Prentice (2017) also used the findings of some students, acquired during the fieldwork. These findings are shown in table 4.1.

Table 4.1: layer dip angle, direction and thickness estimations including averages and standard deviations of 4 fieldwork teams.

	Team1	Team2	Team3	Team4	Mean	Std. Dev.
Dip (deg)	62	80	60	60	*66*	*9.7*
Dip Direction (deg)	160	154	160	145	*155*	*7.1*
Limestone Thickness (cm)	10-50	20-80	5-40	N/A	*12-57*	*7.6 / 20.8*
Marl Thickness (cm)	30-40	20-80	30-300	N/A	*27-140*	*5.8 / 140.0*

These results show numbers which are close to the values found with digital analytics. The dip angle is approximately the same so is the dip direction. The layer thickness estimations are varying much more, which is interesting to notice. These high variation could result from students wrongly identifying the layers or could result from the fact that fieldwork consists out of a small set of samples which is sensitive to incidental measurements.

When quickly compared to the work obtained in this thesis, one sees again that the dip direction is consistent and the dip angle is off by a lot which cannot be compensated for by the standard deviation. The layer thickness calculations made by Prentice is limited by the amount of samples and therefore her

thickness calculations are only representative to a certain extend. The results she obtained fit within the boundaries of the layers obtained in this thesis.

5.3. Comparison with own work

The results from the previously discussed theses are also to be compared with the work produced in this thesis. Comparison between results obtained by different theses is done to validate the results obtained here. This is therefore also the chapter which will answer the following secondary research question

- **How to validate the estimated parameters computed by the computer?**

The validation will be done by comparing the orientation parameters and layer thickness calculations with those of Bisschop (2017), Prentice (2017) and the fieldwork students during the 2016 BSc. Fieldwork trip. Not only will the result be validated by comparing the statistical results, the obtained data is also validated during computation by visually assessing the quality of the fitted planes. This last validation step is not explicitly mentioned in this chapter, it is touched upon in chapter 4.3 where the quality of the fitted planes is assessed

Chapter 4.3 show the dip direction and dip angle of the layers as identified by the designed algorithm. The average dip direction for the layers is 155.96 degrees with respect to north with a dip angle of 51.89 degrees with respect to horizontal. The standard deviations are respectively 3.08 degrees and 0.90 degrees for the dip direction and dip angle.

When you summarize this information and compare it to the information acquired by the theses of Bisschop (2017) and Prentice (2017) and the fieldwork results, one can see the following comparison table.

Table 4.4: Plane orientation calculations of Bisschop (2017), Prentice (2017) and Roebroeks (2018)

Timo Bisschop (2017)	Average	Std. deviation
Dip direction (degrees North)	149.01	30.69
Dip angle (degrees)	66.08	6.25
Elizabeth Prentice (2017)		
Dip direction (degrees North)	157	1.4
Dip angle (degrees)	64	0.6
Joppe Roebroeks attempt 1 (2018)		
Dip direction (degrees North)	156.68	2.45
Dip angle (degrees)	52.38	0.76
Joppe Roebroeks attempt 2 (2018)		
Dip direction (degrees North	155.96	3.08
Dip angle (degrees)	51.89	0.90

Table 4.5: Fieldwork results

	Team1	Team2	Team3	Team4	Mean	Std. Dev.
Dip (deg)	62	80	60	60	*66*	*9.7*
Dip Direction (deg)	160	154	160	145	*155*	*7.1*
Limestone Thickness (cm)	10-50	20-80	5-40	N/A	*12-57*	*7.6 / 20.8*
Marl Thickness (cm)	30-40	20-80	30-300	N/A	*27-140*	*5.8 / 140.0*

One can see from table 4.4 and 4.5 that the results are remarkably close together with respect to one another. The dip directions of the layers are in general very much alike. However, the results of the dip angles obtained in this research differ significantly compared to the results obtained by Elizabeth, Timo and the fieldwork teams.

The large difference in dip angle might come from the fact that the method in this thesis tries to fit planes through the whole layers and not parts of the layers as in all the other theses. In the thesis of Prentice (2017), only a small part of 2x2 meters have been analyzed and the program (LIME) used by Bisschop (2017) also only analyzes a selected part of the outcrop. The student results only include a couple of measurements. This thesis makes use of the whole outcrop and most of its layers and because planes are fitted through whole layers there might be some small deviations with respect to earlier obtained results.

Also the layer thickness can be compared when assessing the different theses and this research.

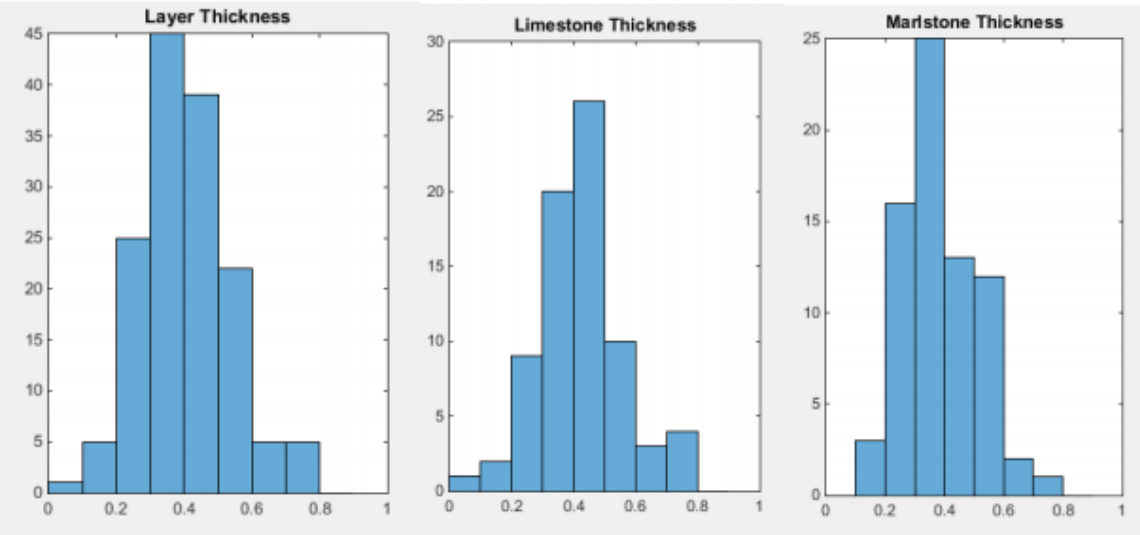


Figure 42: histogram of layer thickness (in meters) by Bisschop (2017).

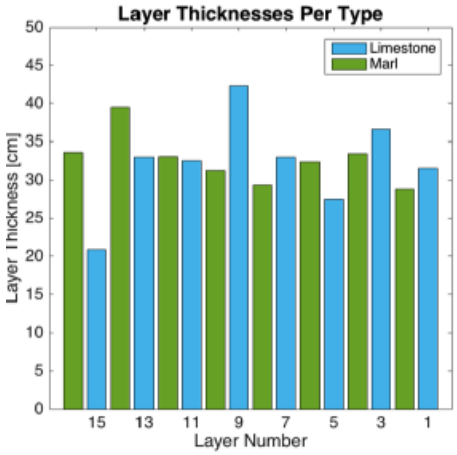


Figure 43: Thickness of layers by Prentice (2017)

Table 4.4 and 4.5 illustrates the layer thickness calculations made by resp. Bisschop (2017), Prentice (2017) and the fieldwork students. These results show very similar results with respect to layer thickness values for both Limestone and Marl layers. In both theses the average layer thickness of the limestone and marlstone is about 0.3-0.4 meters, this is deducted from figure 42 and 43. Table 4.5 shows the same values but with a generally larger spread in the maximum values.

This research also analyzed the thickness of the geological layers. Unfortunately there was no distinction made between marlstone and limestone. All layers have been thrown on the same pile and are evaluated as a whole. The following histogram follows from this computation.

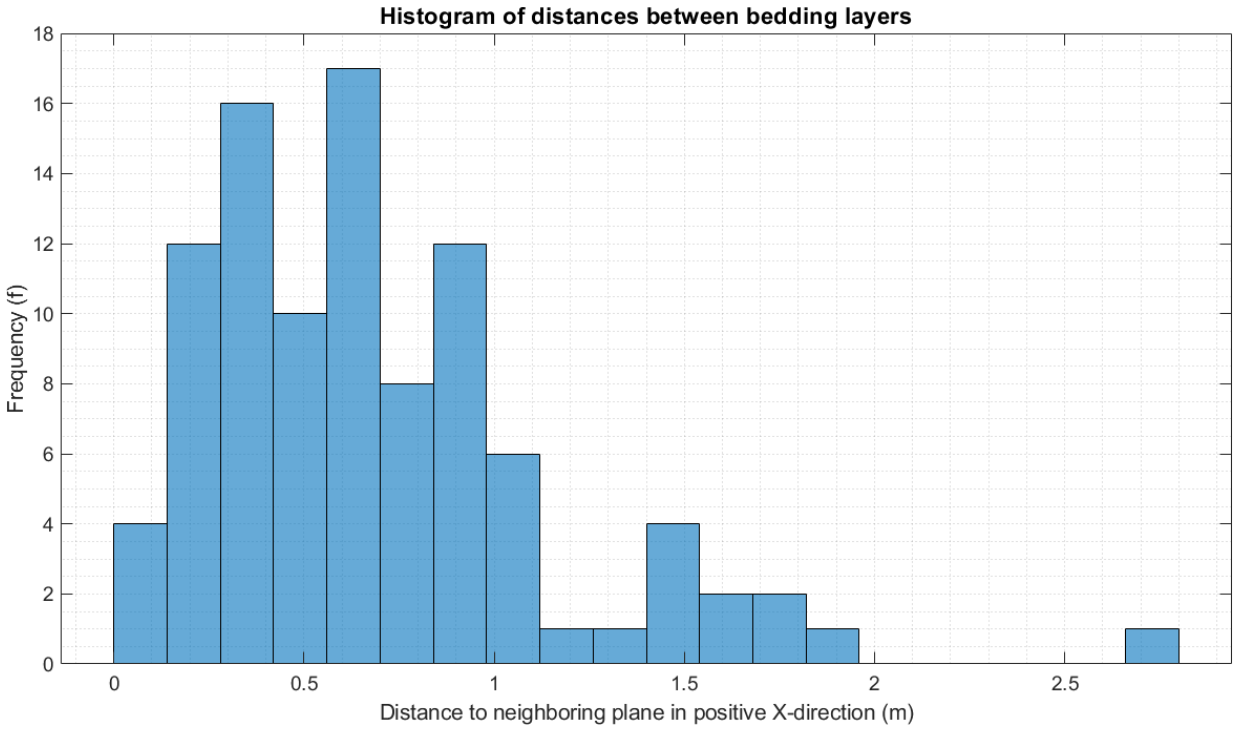


Figure 44: Histogram of computed layer thickness.

When comparing the thickness as computed during this thesis (figure 44) with previous work, one notices that there are similarities and differences. The similarity between the works is that the histogram in the lower region looks very much alike. There are peaks in the layer thickness around 35 centimeters which is supported by all three theses by prentice (2017), Bisschop (2017) and Roebroeks (2018) and around 50-60 centimeters which are only confirmed by Bisschop (2017) and Roebroeks (2018). The fact that Prentice (2017) does not fully support the results found at the other theses comes from the analysis. Elizabeth Prentice only evaluated 2x2 meters while the other theses did the whole outcrop.

Summary.

The work performed by Prentice (2017), Bisschop (2017) and this thesis is consistent with exception of some details. Layer thickness calculations are very much alike as well as the orientation parameters. However, more work has to be done on the code written during this thesis in order to make it better recognize planes and to better prevent commission and omission errors.

5.4. Parameter settings

This method is not automatic, it is even far from that. A part of the parameters is chosen by hand and supported by the work of Prentice (2017). Some parameters are automated in order to be able to use this code on other geological outcrops. As noted before, this method contains a graphical user interface so that users have control over what data is actually worked with. This control can turn out wrong if not correctly used. However if this control is correctly used, it will be a very powerful tool during processing.

The method relies on a vast series of separate computations with several functions. Below will be a description of the settings and parameters used during the whole process.

5.4.1. Subsampling

The first step in processing is equalizing the data density. Therefore a subsampling step is performed on the whole dataset. This most equal data density is acquired by sampling with equal distance. The parameter which is considered is the following:

- **Mimimum space between points.** This is the minimum space between two arbitrary points in the dataset, it has to be a minimum of the provided distance.

The minimum distance between two points in the point cloud has chosen to be at least 1 cm or 0.01 meters. Often the outcrops used are heavily scanned, resulting in some parts more densely scanned than others. Data density can be higher than 1 point per cubic centimeter, this parameter is merely to make the dataset more uniform.

5.4.2. Roughness calculation

This processing step is necessary for later filtering. The roughness is based on the following parameter:

- **Kernel size.** This is the minimum search radius within points are grouped for plane fitting.

The kernel size has chosen to be 0.4 which stands for 0.4 meters. Choosing this value for the kernel size shows great differences in roughness in the features in the “La Charce” outcrop. They can presumable also be used on other outcrops.

5.4.3. Normal calculations and orientations.

Normals are the most important piece of information created and used in this thesis. There are several parameters involved in the process. These are the following:

- **Amount of neighbors.** The amount of nearest neighbors (6) are used for local plane fitting to determine a point normal. The greater the number the smoother the normals become.
- **Vertical vector.** The vertical is used to again re-orientate normals upwards in order to coincide with plane orientation conventions.

As mentioned, the more neighbors are used, the smoother the distribution of the normals will be. However, a smooth distribution of normals is not preferable since it provides no local information on the points. A **nearest neighbor count of 6** has been used since it is standard for Matlab, it provides normals with more than sufficient information on the orientation of the points with respect to their neighbors.

For the vertical vector, this is just a vector pointing straight up (which is the vector [0 0 1] when considering coding language). As mentioned before this parameter is used to make normals of points representable as normals of geological planes and can be considered as such. When geological outcrop

are evaluated by hand, the dip direction and angle of a layer is conventionally evaluated with respect to the top of the bedding. The property used in this case is that the normal of this bedding is pointing upwards (or downwards) and therefore this upwards pointing property is very well usable for further computations.

5.4.4. GUI preparations

The graphical user interface requires some settings and some filtering before it can be used.

- **Subsampled point cloud size.** This is the size of the subsampled point cloud which is going to be used by the user. The larger this subsampled point cloud, the better the user will be able to accurately select an area of interest.
- **Max roughness changes.** The default maximum roughness can be very high and only few values are included. The top part of the roughness histogram can be rescaled in order to prevent the outliers in the roughness dominating the visualization of the roughness values.

The point cloud size has been chosen to be 200.000 points. This amount of points is enough to visualize features in the outcrop such as planes, corners, rough edges and discontinuities. This value can be higher (maybe 400.000), which will enhance the quality of the used GUI figure. Unfortunately it will also increase computation time. The more points are included in the figure, the slower the point acquisition will be when the user selects a point in the figure. Fortunately the speed at which points are stored while at 200.000 points is very fast, and thus higher amounts of points will most likely not be a bottleneck for the user experience.

Roughness values are altered in order for the above produced figure to be pleasantly visualized. The highest 10 percent of the values are changed to the value located at the 90 percentile. This 90 percentile has been chosen based on the shape of the histogram of the roughness values. The distribution looks like chi-squared or gamma distribution. Therefore there are relatively small amounts of points with very high values dominating the edge of the distribution. The highest 10 percent of the values have been cut off and replaced by the value found at the 90^{th} percentile.

5.4.5. GUI result bedding plane extraction

After using the graphical user interface, selecting points in the figure, computation continues with extracting small clouds of points based on the selected points. These points are then subjected to an intensity and roughness filter and the remainder of those two filters is subjected to a histogram selection stage. The dominant orientation parameters from these 3 filtering steps is used in the plane fitting algorithm, after the dominant orientation has been used to select all relevant points and an additional filtering step is applied to the whole point cloud. All these steps require certain settings. Below is a list of all settings and parameters used for this and their numbers:

- **Search box.** Around the selected points by the GUI, a virtual box is drawn within all points are considered to be included for the computation. The size of this virtual box is 0.4 x 0.4 x 0.4 meters by default, it is about the size of 1 of the dimensions of the features you try to extract (which means in 1 direction the feature is about 0.4 meters in size).
- **Roughness limit.** High roughness values represent rough areas. These are not desired since the interest lies in smooth surfaces. High roughness values are filtered out in an intermediate computation with a limit of 0.01 meters. Everything with a higher roughness than that is excluded, this is approximately the 50^{th} percentile in the data.

- **Intensity limit.** High intensity values represent smooth surfaces with little to discontinuities. This filter will support the roughness filter. The intensity has to be at least 0.3 (30%), everything below this number gets excluded in the intermediate computation (this is about 50%). The roughness and intensity filter can be automated by selecting the 50^{th} highest percentile and performing the same computation.
- **Potential bins.** The amount of bins determine the angular precision of the histogram selection. By default this is 60 bins which brings the angular differentiation to 6 degrees per bin. This is sufficient to identify peaks in dip directions. If the angular differentiation is not enough, the total amount of bins can be reduced to increase the counts per bin (and thus increase the chance to find a dominant direction). However, decreasing the amount of bins will reduce the chance of immediately acquiring the correct dip direction since additional noise is introduced.
- **Number of neighbors denoise.** This number is used for the denoise function as a neighbor count. Default is set to 50 neighbors as that provides adequate diminishing of random noise in the point cloud.
- **Threshold denoise.** The threshold is set to 2.4 centimeter which is the distance between a sample point and its 50 neighbors. If the mean distance is larger than 2.4 centimeters, the point is excluded. 2.4 centimeter is a numbers which proves sufficient elimination of noise. A small analysis will follow later.
- **Minimum points per plane.** A minimum of 2000 points per plane is introduced such that the plane fitting algorithm does not fit (near) infinite amounts of small planes. Even though filtering has been thoroughly done, small clusters of undesired points remain. A minimum of 2000 points per plane should prevent random planes from being fitted most of the times.
- **Maximum distance plane fitting.** The maximum distance between the fitted plane and a point belonging to that plane is set to 15 centimeters (so a total tolerance in both directions of the plane of 30 cm), this because the surface of the outcrop will not be smooth. 30 centimeters proves to be adequate to incorporate geological planes by full (mostly).
- **Angular difference plane fitting.** The maximum angular difference from the provided normal for plane fitting algorithm is set to 5 and 6 degrees (for run 1 and run 2). A small error is expected since the selection stage of the dip direction and angle is not perfect. This 5-6 degrees of freedom should provide for adequate room for any small changes the method decides to make.

All the above mentioned parameters are estimated, investigated and accepted as adequate to prove the use of this method. In chapter 5.5, a small sensitivity analysis is performed in order to assess and show other configurations of parameters.

5.4.6. GUI results computation fold

When performing a fold identification, a part of the methods used is the same, only the configuration for the parameters was slightly different and some steps were even excluded because the computation without was sufficient. However the settings used are explained below.

- **Search box.** The search box for finding relevant points has been significantly reduced since the outcrop evaluated is much smaller. The search box size is reduced to 15 cm in each direction, the box dimension is thus 0.3m x 0.3m x 0.3m in size. This proved more than adequate in order to provide a good result.

- **Roughness limit.** The roughness limit for intermediate filtering has been set to the same value of 0.01 meter. If intensity was present it could be kept at 0.3 just like in the previous calculation.
- **Intensity limit.** No intensity limit was present in the outcrop, and thus no value was determined and used.
- **Estimated dip direction.** The estimated dip direction is provided by the user, I estimated 150 and 100 degrees both either sides of the fold. These angles can be off but will be corrected by the method. The tolerance of 50 degrees will resolve that issue.
- **Number of neighbors denoise.** The outcrop used has a much higher point density compared to the other outcrop. Therefore, stronger limitations with respect to filtering need to be maintained. The number of neighbors has been set to 500 which results in only large sections remaining.
- **Threshold denoise.** The denoise threshold stayed 2.4 centimeter.
- **Minimum points per plane.** The minimum amount of points per plane has been decreased from 2000 to 1000 points for this case since the outcrop is much smaller.
- **Maximum distance plane fitting.** The maximum distance from the plane model has been kept on 15 centimeter.
- **Angular difference plane fitting.** The angular difference which the plane fitting algorithm can introduce is equal to 5 degrees.

5.5. Sensitivity analysis.

Parameters chosen as in chapter 4.1.1 can vary according to preferences. However, the parameters chosen for this research are considered to be satisfying and sufficient for the goal of this research. Theoretically if for every step 3 possibilities would be considered (3 variations on the possible parameters), the sensitivity analysis would consist of 3^12 to 3^14 different possible outcomes. These numbers are far too great to analyze. Therefore, for only some of the parameters (if they are considered of major influence) a sensitivity analysis is performed.

When the sensitivity of 1 parameter has been visualized, the optimal parameter mentioned in paragraph 4.1.1 will be used in the next step. This way, only a maximum of 42 results have to be analyzed. The end result of any of the computations is not visualized, only the intermediate results are shown.
The investigated parameters are:

- Kernel size for roughness calculations

For both normal plane fitting and fold identification:

- Roughness limit
- Intensity limit
- Number of neighbors denoise
- Threshold denoise
- Minimum points per plane
- Maximum distance plane fitting
- Angular difference plane fitting

These are considered to be the most important parameters, resulting in a sensitivity analysis where 42 intermediate results are evaluated. These results are grouped in pictures of 3 and will be briefly explained. One of the results shown in every subpart of the sensitivity analysis is the original used in the research results.

Kernel size

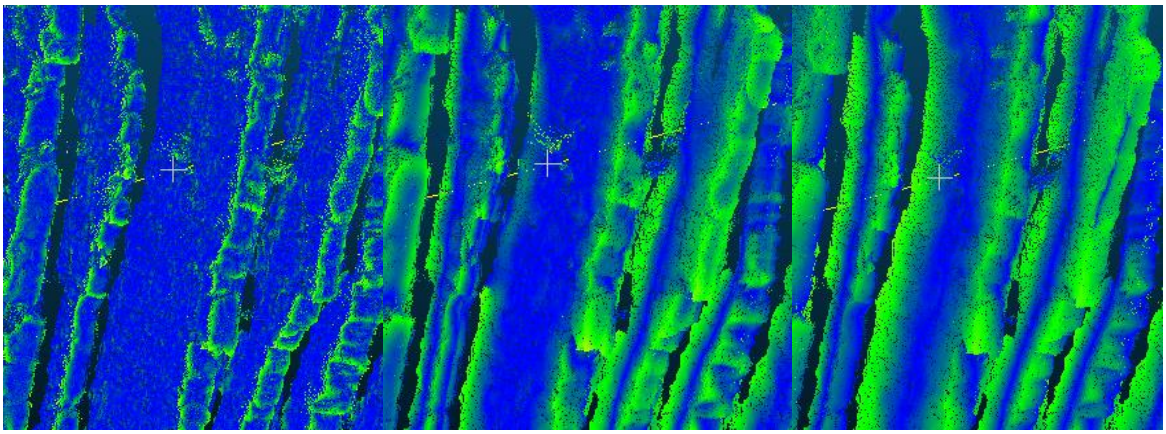


Figure 45: Kernel size of 0.1m, 0.4m and 0.7 meters.

In figure 45 one can see three images computed with Cloudcompare. The images are a visualization of the roughness index which has been given to every pixel. In the most left image one can see that at the edge of the geological layer are indicated with bright green. Bright green indicates that there is a higher roughness value, approximately 1 centimeter or higher (1cm is the threshold which is used as a filter later on). The 0.1 meter kernel is therefore considered to be too little to filter out crucial values.
The kernel size of 0.4 and 0.7 however, show similar results with respect to desired and undesired parts of the geological outcrop. The 0.7 kernel size does require a lot more computation time while it does not show and significant improvement over the 0.4 kernel size computation. Therefore the 0.4 kernel size computation is considered to give the best result and is used for the next step.

5.5.1. Bedding plane extraction

Roughness limit.

The roughness limit is used to filter the points which have been previously selected by the Graphical User Interface. This filter is together with the intensity limit looked at later onwards, this filter exists to prevent undesired points from being included in the computation. The sensitivity analysis has been performed on three different roughness limits given below:
- 0.005 meter (much smaller than the used roughness limit).
- 0.01 meter (the used roughness limit for filtering).
- 0.03 meter (chosen to see what a larger roughness limit will result in).

The result from this is a different average normal vector, the average normal vector is given below in the same order.
Normal 1: [-0.8316, -0.3718, 0.3552].
Normal 2: [-0.8310, -0.3720, 0.3571].
Normal 3: [-0.8298, -0.3724, 0.3546].

The resulting 3 normals differ mostly in the 3rd decimal, resulting in the differences between the angles being approximately 0.1 degree (following from calculation). This shows that within reasonable limits of the available roughness values, changing the roughness limit will not significantly affect the used normal for plane fitting.

Intensity limit.

Just like the roughness limit, the intensity limit also influences the points picked in the GUI for the same purpose. The sensitivity analysis has been performed on 3 different values.

- Intensity limit of 0.1 (a lot smaller than the used value for this thesis).
- Intensity limit of 0.3 (this is the used value).
- Intensity limit of 0.4.

The resulting average normals are shown below in the same order

Normal 1: [-0.8250, -0.3675, 0.3539].

Normal 2: [-0.8310, -0.3720, 0.3571].

Normal 3: [-0.8192, -0.4163, 0.3126].

The angle between normal 1 and normal 2 is not significant, only 0.1 degree. However, the angle between normal 2 and normal 3 changes the angle of the average normal by 3.7 degrees. This means that the intensity has a more significant impact on the resulting normal used for plane fitting. However, the chosen intensity limit 0.4 is so exclusive that there are barely any points remaining when it is used. Therefore it can be concluded that this value is also barely affecting the normal vector used for plane fitting.

Number of neighbors for noise reduction.

The noise reduction algorithm by Matlab will reduce the noise after roughness and intensity filtering is performed. The values chosen for the nearest neighbor noise reduction will have an effect on the remaining points. The used value in the report is 50 neighbors, and therefore values of 10 and 100 have been evaluated in order to see what happens in those cases. The result is shown in figure 46.

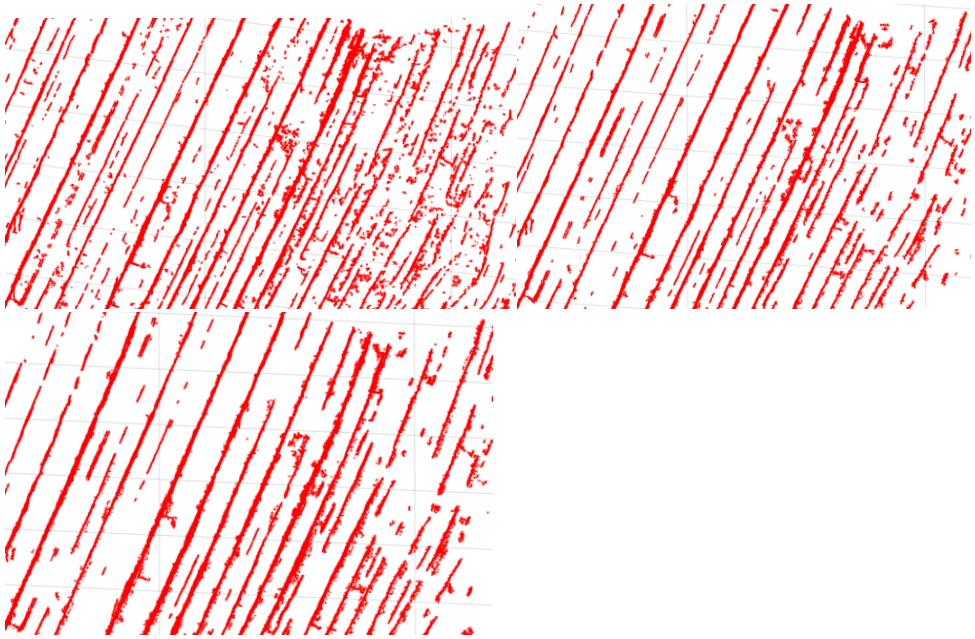


Figure 46: 10, 50 and 100 neighbors for noise reduction algorithm

One can see from figure 46 that there is a significant difference between the point clouds as they are being filtered with different values. One can see that in the case of 10 neighbors, a lot of noise remains together with some small layers. In the case of 50 neighbors which has been used for this thesis, the layers are much more divided and much clearer. A lot of noise in between the layers has been reduced although that also comes with the cost of filtering (removing) some geological layers. When using the 100 nearest neighbors, almost all the noise is gone but unfortunately also all the layers smaller and lesser sampled layers disappear.
Unfortunately these noise reduction algorithms cannot see what is supposed to be layer, and thus layers will always be filtered out in some way since they just fall within some numbers. As long as no distinction can be made with respect to certain areas or structures, some consideration has to be made. Noise will be reduced but at cost of some layers as well.

Threshold denoise.

The threshold used for the noise filtering algorithm in Matlab has been evaluated for three different values (of which one is the used value in the thesis). A very large range of the threshold has been chosen in order to see its effect. Unfortunately the difference wasn't really noticeable in the resulting point cloud. Therefore only the size of the point cloud has been noted down.

- Threshold of 0.1 meter results point cloud size of 1465018 points
- Threshold of 0.024 meter results point cloud size of 1342025 points
- Threshold of 0.0001 meter results point cloud size of 1292182 points

The images belonging to this evaluation can be found in appendix 2 in figure 1. When evaluating the difference in the amount of points compared to the threshold numbers, one sees that a difference in the threshold of a factor of 1000 only reduced the amount of points by approximately 12 percent which raises questions with respect to the influence and

implementation of this parameter. One can therefore not make a specific conclusion with respect to the influence of the noise threshold else than saying that it works.

Minimum points per plane.
The plane fitting algorithm for the normal cases has a set of parameters which can greatly influence the results. One of those parameters is the minimum plane size. The higher the minimum plane size, the potentially less planes are fitted through the dataset. The size of an outcrop determines the minimum size of the planes since the maximum point density is going to be the same in all outcrops (because of subsampling). The 3 evaluated situations are the following
- Minimum size of 1000 points resulting in 113 planes.
- Minimum size of 2000 points resulting in 106 planes.
- Minimum size of 5000 points resulting in 61 planes.

The images belonging to these computations can be found in Appendix 2 in figure 2, the images are in order. One can clearly conclude that a minimum plane size significantly influences the amount of fitted planes. This is the case because there is a limit of the amount of points belonging to a bedding plane. As soon as a bedding plane has less than the set amount of points, it will not be defined as such. Setting the minimum amount of points per plane to 2000 for this specific outcrop yields the best result with respect to bedding plane identification.

Maximum distance plane fitting.
The maximum distance to the plane influences the results as well, but not in the way which one would expect. Three different situations are evaluated as mentioned below, together with that the amount of planes identified will be mentioned.
- Maximum distance of 7 centimeters yields 40 identified planes.
- Maximum distance of 15 centimeters yields 110 identified planes.
- Maximum distance of 25 centimeters yields 84 identified planes.

These results might not look as expected. The reason for this is because the maximum distance behaves based on the bedding thickness of the geological layers and the spread of the points belonging to 1 bedding plane.

Choosing a distance of 7 centimeters will yield only 40 planes since not enough points can be assigned to 1 specific plane accommodating the minimum plane size of 2000 points. The model is possibly also performing omission errors since the constraints are too high. In the case of a maximum distance of 25 centimeter (84 planes), commission errors will occur. Since the maximum distance to the plane is 25 centimeters, a total spread of 50 centimeters can be included in the results. Since the bedding thickness is often less than 50 centimeters, different geological layers can be fused together since the model allows it. 15 centimeters yields the best identification of the planes since most layers have a thickness of less than 30 centimeters and thus no commission errors and only few omission errors are made. The fitted planes are shown in Appendix 2 at figure 3.

Angular difference plane fitting.
The angular tolerance, the angular freedom of the plane fitting, has been varied in order to see the influence of the angular freedom on the obtained results.
- Maximum angular tolerance of 1 degree yields 57 planes.
- Maximum angular tolerance of 5 degrees yields 107 planes.
- Maximum angular tolerance of 10 degrees yields 96 planes.

The images belonging to this computation can be found in Appendix 2 at figure 4. The angular freedom of 1 degree is obviously very little, the given normal for plane fitting might not be entirely correct and therefore the angular tolerance is needed. And angular tolerance of 1 degree is too little to correct for possible mistakes in the given normal. 5 and 10 degrees of tolerance yield much better results with respect to fitted planes and the correctness of the fitted planes. The only issue is that when the tolerance is too large, commission errors will occur again and this is exactly what is happening when a tolerance of 10 degrees is provided. Therefore an angular tolerance of a few degrees (5 in the case of this case and most other datasets), is sufficient to correct for possible mistakes in the provided normal.

5.5.2. Fold identification
For the fold identification the roughness filter has been varied according to the values present in the data. The standard roughness filter filters out any value above 0.01. For comparison, filter values of 0.003 and 0.02 has been used to see what difference it makes for the estimated normals on both sides of the fold. The results are shown below.

Side	**Roughness limit**	**U Component**	**W Component**	**V Component**
Side 1	0.003	-0.4343	-0.5418	0.7080
	0.01	-0.4375	-0.5436	0.7018
	0.02	-0.4385	-0.5448	0.6989
Side 2	0.003	0.0749	0.9519	0.2708
	0.01	0.0726	0.9511	0.2711
	0.02	0.0723	0.9508	0.2716

One can see from these results that the roughness limit barely makes any difference with respect to the estimated normal. A factor 6 shift in the roughness limit only causes a change in the normal which is visible in the 3rd decimal of any of the components. This change is so small that it only accommodates for a normal change of 1 degree or less. Again the roughness filter does not cause great differences.

Number of neighbors denoise.
The number of neighbors for this dataset do cause great differences with respect to present noise in the outcrop. The standard amount of neighbors has been set to 500, this to make sure that only big planes are kept. As one can see in the images below, the difference between 50 and 100 neighbors is not very well visible. 500 neighbors does cause a great difference in comparison to 50 and 100 neighbors. The 500 neighbors result removes much more small planes (which might not be desired) but also makes sure that only large clusters of points, which are geological planes, remain. The colors blue and red stand for

either side of the fold. Blue is for side 1 of the fold and red is for side 2 of the fold. Side 1 and side 2 can of course be chosen arbitrarily. The results of the different neighbor noise reduction trials are illustrated in figure 47.

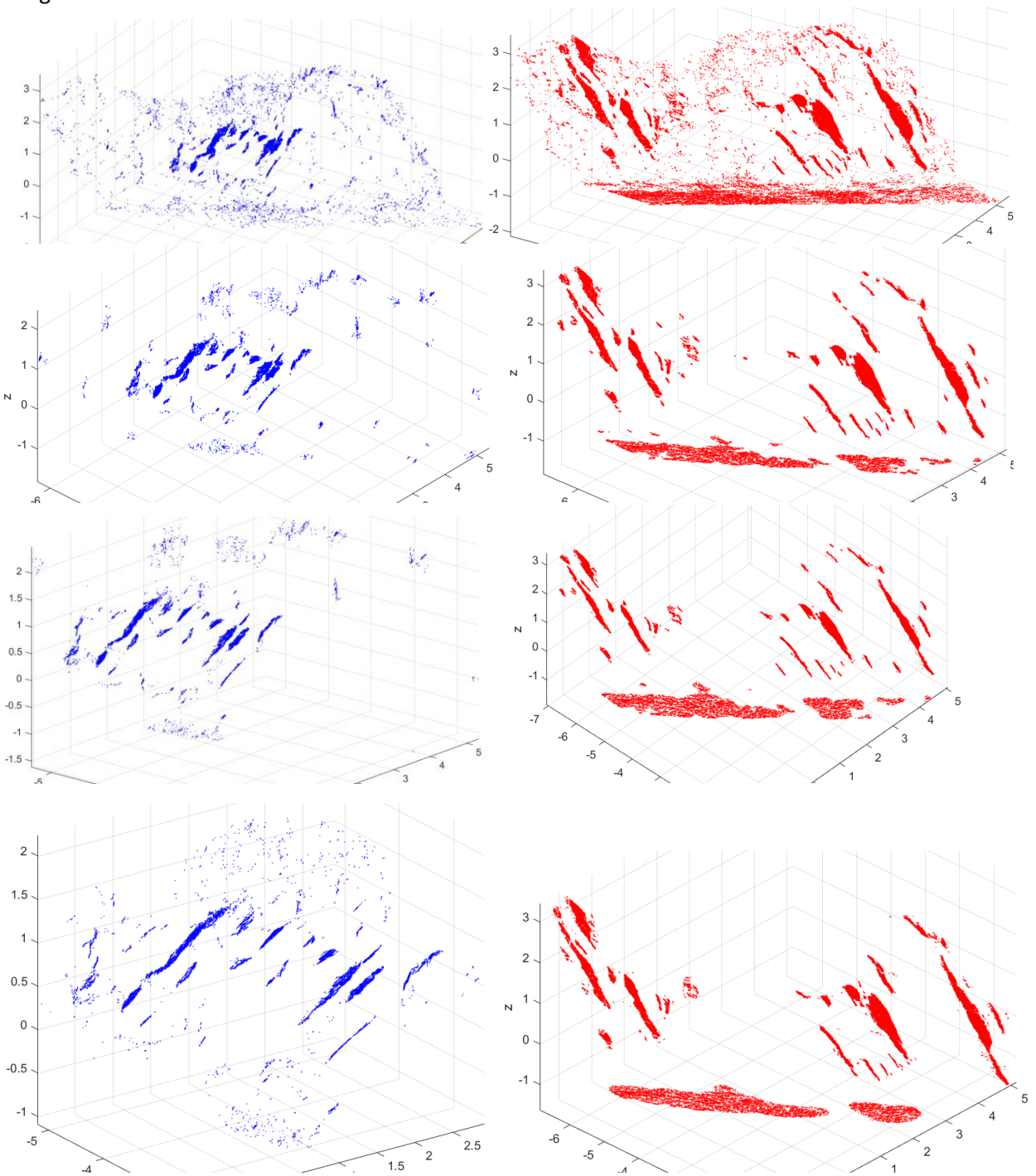


Figure 47: Horizontal downwards are shown the original pointcloud selection (with directional filter and roughness filter), the 50 neighbors result, the 100 neighbors result and the 500 neighbors result

Unfortunately, choosing a higher number also increases computation time, choosing 500 neighbors causes a 15 minutes computation time. But on the other side it generates the most favorable result possible.

Threshold denoise.

The second parameters used for the denoise algorithm is the threshold beyond which points are excluded. The used threshold is 0.024 meters with variations from 10 centimeters to 0.1 millimeter. The result is not very well visible in the resulting point clouds even though there is a difference of a factor 1000. Since there is no clear visible difference, the amount of points will be mentioned in the table below.

Threshold (meters)	Side 1 (amount of points in PC)	Side 2 (amount of points in PC)
0.1	7986	37348
0.024	7706	36214
0.0001	7611	35832

The maximum difference between the 0.1 and the 0.0001 meter threshold is only 5%, there is still a noticable difference present. The smaller the threshold becomes, the less points are left in the remaining point cloud. Unfortunately the difference is not as much as expected.

Minimum points per plane.

The minimum amount of points per plane in the fold dataset has also been evaluated. The default size in this case is 1000 points, 2 variations of 500 and 2000 points have been made in order to see the difference. The images belonging to this computation can be found at appendix 2 in figure 5. The resulting amount of planes are summarized in the table below.

Minimum amount of points per plane	**Amount of identified planes side 1**	**Amount of identified planes side 2**
500	6	8
1000	2	10
2000	1	4

One can conclude from this table that the larger the minimum plane size, the less planes will be identified. This is caused by the fact that this outcrop is significantly smaller than the La Charce outcrop. The minimum plane size at the La Charce dataset is 2000 planes, it has been decreased for this outcrop since be default less points will be present in a plane even if it is present since the whole outcrop is smaller. It is concluded that 1000 points per plane is the best choice, resulting in a minimal amount of commission errors.

Maximum distance plane fitting.

The maximum distance to the plane has also been evaluated in the sensitivity analysis. In general: the smaller the maximum distance, the smaller the planes and the less planes will be identified. The default maximum distance has been set to 15 centimeters because of an expected layer thickness not exceeding 30 centimeters on average. The images can be found in Appendix 2 in figure 6. The summary of the amount of planes is shown in the table below.

Distance (centimeters)	Amount of identified planes side 1	Amount of identified planes side 2
7	**2**	**8**
15	**2**	**10**
25	**2**	**9**

Based on this information it has again been concluded that a maximum of 15 centimeters results in the best possible configuration for this specific outcrop. 15 centimeters is just like in 5.4.1 a good choice for the maximum distance to the fitted plane. The influence of the maximum distance depends on the quality of the outcrop and the distance over which the bedding is visible. The larger the bedding size, the more the maximum distance will be of influence. This is also evident from 5.4.1 since a maximum distance of 7 centimeters resulted in only 40 planes while a larger distance resulted in a factor 2-3 more planes.

Angular difference plane fitting.

The last parameter which has been evaluated is the maximum angular difference. The default angular tolerance is 5 degrees and the variations are 1 degree and 10 degrees and the results are visualized in appendix 2 figure 7. The summary of the amount of planes identified is shown below.

Maximum angular tolerance Degrees	Amount of identified planes side 1	Amount of identified planes side 2
1	2	10
5	2	10
10	2	8

Again the angular tolerance does not greatly influence the amount of identified planes. It does however influences how the plane fitting algorithm fits the planes through the available datasets. A lower tolerance results in more parallel planes which is logic. The tolerance cannot be too large in this case, since a large tolerance will result in commission errors where planes are misidentified.

6 Synthesis

This chapter provides more discussion material and an elaboration on some additional computations performed together with the viability of using this method on other datasets.

6.1. Results

The results obtained by this research show mixed results with respect to progress. A lot of progress has been made with respect to data management. This method has been tested with 50 times more data than the predecessor used. The results show more progression with respect to automation and streamlining of the whole process. Positive results with respect to calculation time are also obtained. Calculation time has been decreased compared to previous calculation methods down from approximately 1 hour to a few minutes depending on the operating system running the script.

The results obtained are more satisfying with respect to automation, but less satisfying with respect to progress. The previous theses had an identification of layers where this thesis does not have that distinction.

6.2. Applicability method.

The script created during this thesis is designed to be compatible with other datasets just like software packages are designed for other datasets as well. Programs such as Lime, Cloudcompare and Meshlab provide some functionality with respect to geological outcrop analysis. Unfortunately, those programs are not made to analyze a geological outcrop in full. Only parts of the script functionality created during this thesis are available in other free available software. Apart from functionality, the script is also faster than software available online. Computing normals takes very long within Cloudcompare when one choses to process large point clouds and Meshlab is not able to efficiently process more than 10 million points at the time. These are all program limitations which are not present in this script (up to a certain extend). Computation and temporary storage is relatively fast within the created script. Parallel computing is used where possible and profitable and the limitation with respect to how much points are stored depends on your system memory. Matlab uses system memory less efficient than most other point cloud computing programs, but it is able to process much faster in a semi-automated way making it more appealing to use. LIME is a geology specific point cloud processing tool which competes very well with the created script. Lime is better with respect to accuracy, unfortunately LIME is very labor intensive and creates only 1 result at the time (identifies layers and their thickness in the case of Bisschop 2017). If the functionality of LIME improves, it could become more interesting and more competing with home-made scripts.

The reason why this script is more appealing than any other of the programs is that it is faster to use than any other product. It is a script which is more specific and generates more answers to geological questions than any other method can do. In roughly 15 minutes (give or take a few minutes depending on system capabilities), any of the geological features mentioned in chapter 2.1.3 are obtained. This because parallelization of cores is used, code is written efficiently and data storage/access is faster than most other programs.

The script is designed to be compatible with other datasets. This compatibility is only true when datasets fed to the method are built in the same way as the dataset used to create this thesis, other

requirements with respect to the outcrop are mentioned in chapter 3.4. All these requirements have to be fulfilled in order to be able to use this script on other datasets. Of course this would not be proven if this method would not be applied to other available datasets. Therefore an analysis is done with the same script on another dataset.

The dataset origins from the same fieldwork as the La Charce dataset and is therefore located roughly at the same location in france, the precise location is shown in figure 48 and the dataset is called the Pradelle 2 dataset.

Figure 48: illustraion of location of dataset near the village of Pradelle

When opening the dataset with Cloudcompare shown in figure 49, one can directly see that this outcrop is not of the same quality as that of La Charce, This outcrop contains much more noise data such as vegetation and rubble. This makes computation less accurate and less relevant but it will be done either way. Point cloud properties are the following:

- Number of points: 37,646,753
- Point density: 10 points per cm^3
- Number of images or scans: 3 scans are made in high resolution
- Features of scans: position parameters X,Y,Z, colors RGB and intensity I.

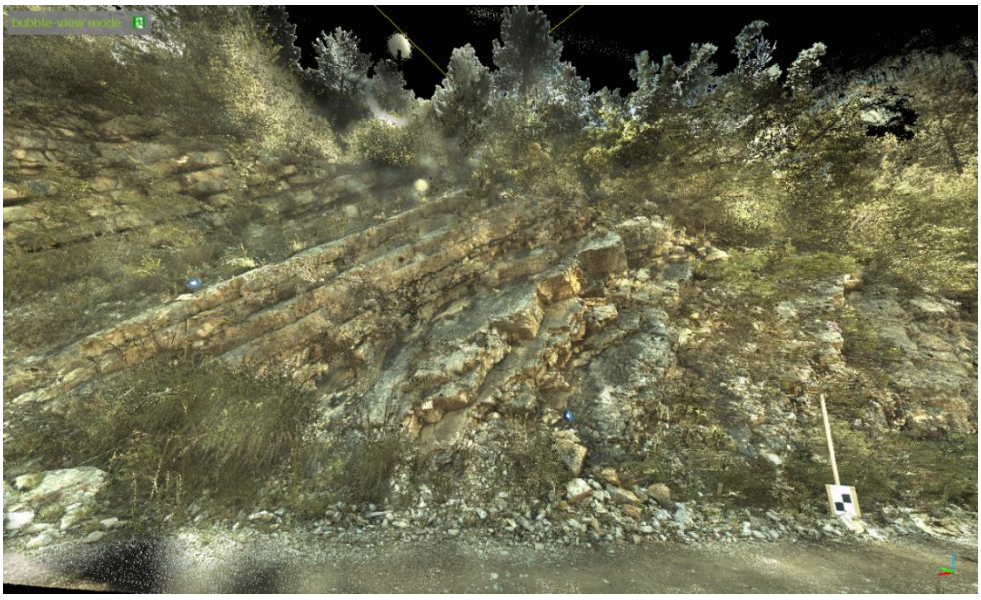


Figure 49: impression view of the Pradelle 2 dataset.

When the script is done processing the Pradelle 2 dataset, it does identify what it should identify: bedding layers. Most of the bedding layers shown in figure 49 are identified by the script, this identification is shown in figure 50 but it is not very well visible because the results look slightly messier with noise datasets. Noise reduction is therefore also something which could be done better in future work with this data. Noise is the major problem for identifying the planes, commission and omission errors are still common errors made and the focus should be on reducing those errors in future work with these types of datasets.

All together it can be concluded that this script is usable on other datasets if the conditions in chapter 3.4 are met together with dataset compatibilities (datasets need to be built in the same way).

An important question for geologists is now:

- **How can this method help geologists in assessing geological outcrops?**

In other words: what is the added value to geologists? This script adds value to geologists because it provides the possibility to evaluate the whole outcrop rather than just some parts. Geologists are in general good in analyzing an outcrop according to some measurements, but this method can help the geologist to evaluate parts which the geologist can't reach. The second power of the digital outcrop version is that the geologist can confirm his findings in the field by looking at the results of the data.
The script also helps in defining layer thicknesses and can in the future (with some more extension) be used to make whole 3D geological models according to data obtained my laser scanners and processed by scripts such as these.

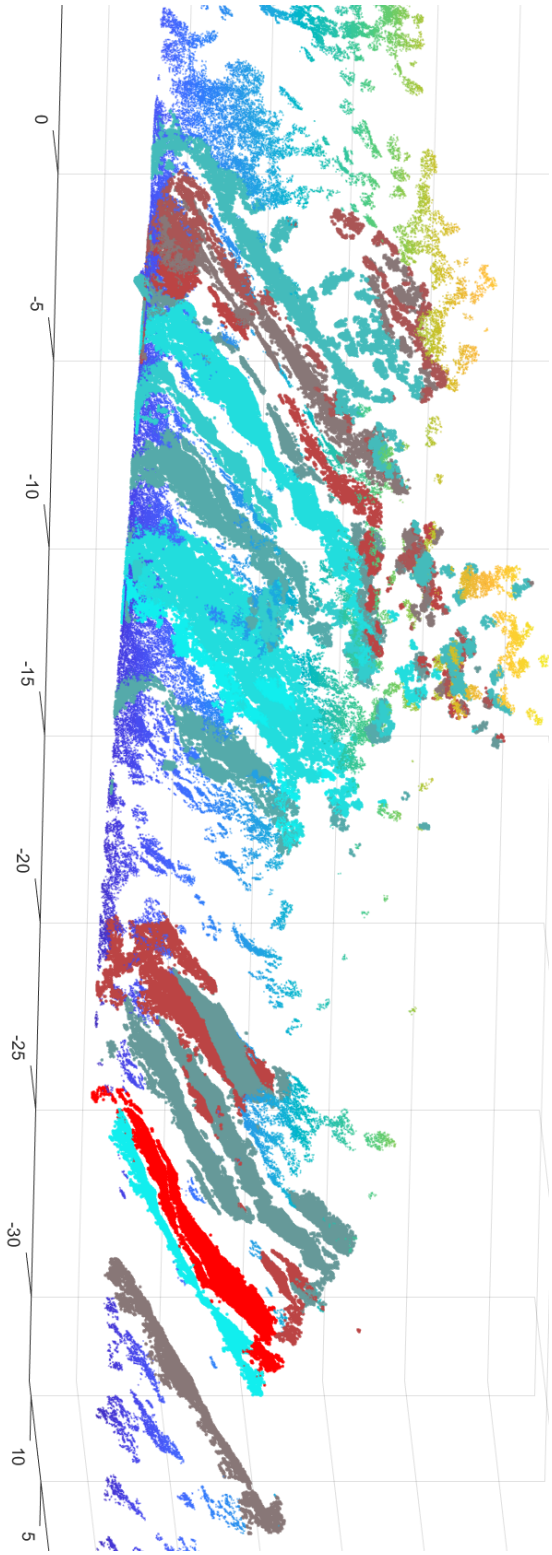


Figure 50: The result of the Pradelle 2 plane fitting

7 Conclusions and recommendations

During the course of this thesis, an extensive look has been taken at how to obtain more geological information from large 3D point cloud datasets obtained from geological outcrops. The datasets included are the famous La Charce outcrop near the village of La Charce illustrated in figure 51 and a dataset of a fold visualized in figure 20. The script made during this thesis was meant to be as autonomous as possible, which it became more up to a certain extend as the script evolved. This study will conclude with comprehensive answers on the research questions, followed by an elaboration on the goals set in chapter 1.1 and closing with recommendations for further research.

Figure 51: La Charce outcrop which is closely accessible and a very good example to test the created script.

7.1. Addressing research questions.

The research questions have been mentioned all throughout the text, a conclusion will follow where the research questions will be addressed again and a comprehensive answer will be given to each of them.

- **What geometric parameters should be extracted from point clouds?**

The geometry parameters of a geological outcrop are its features with respect to geology. The following parameters will be extracted from the large point cloud datasets:

- Layer orientation parameters of parallel geological layers.
- An indication of a potential curvature in parallel geological layers.
- Layer thickness (also called strata thickness) of the parallel geological layers.
- Orientation of geological layers on either sides of a fold.
- Axial surface of a present fold.

- **What existing point cloud processing methods could be useful for such feature extraction?**

There are software tools such as LIME, Cloudcompare, Meshlab and Matlab. Each tool has its specific point cloud implementation and they can all be used for the purpose of this thesis in some way. Unfortunately, using more programs also causes more complications and therefore only Cloudcompare and Matlab are used.

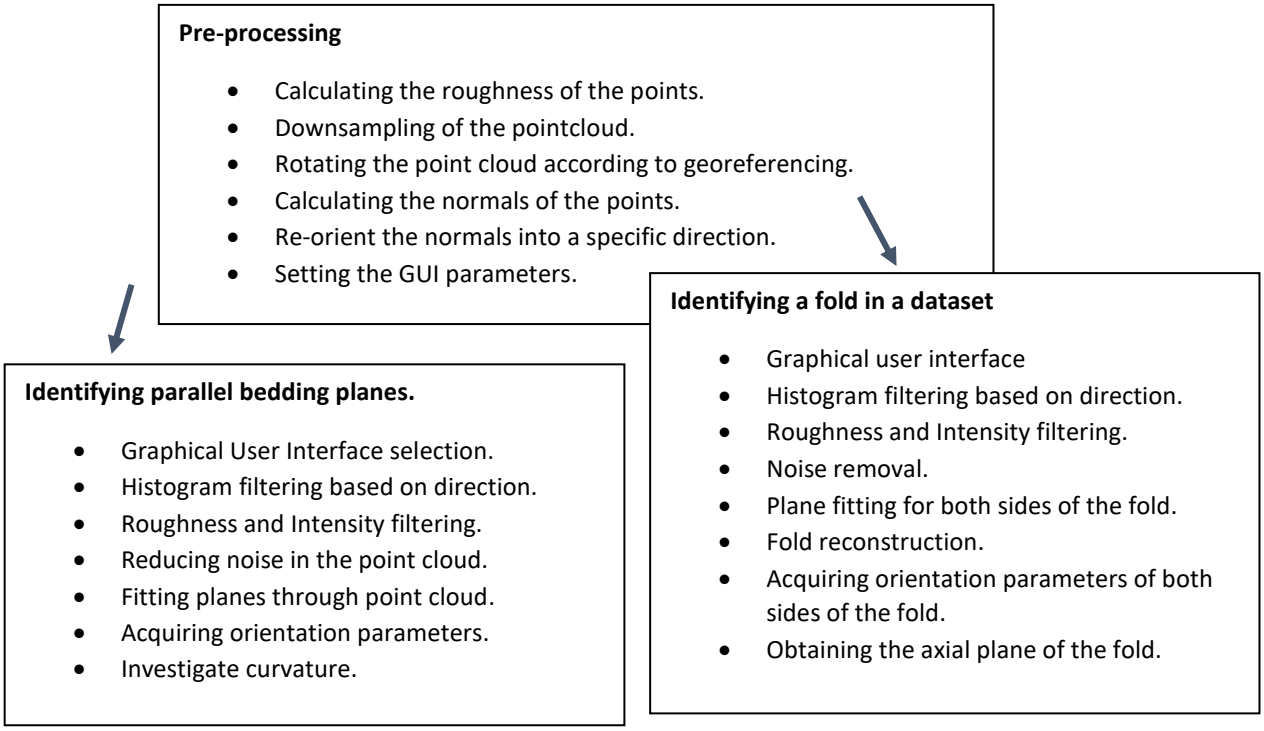
There is also existing work done by researchers. Their work included for example Intensity corrections for distance and angle complications, filtering based on local densities and a Label Connected Components method separating clusters of points in a point cloud. There are also software implemented methods such as noise reduction algorithms, roughness calculations, RANSAC and normal calculation methods for points. For further elaboration on the above mentioned work, please look at chapter 2.4 and 3.2.

- **What are the characteristics of laser scanning point clouds?**

The main characteristics of Laser scanning point clouds is that datasets are very big, often 10's of millions of points are included in the dataset which makes computation a difficult task. Computations are time consuming if data is not handled with care. Secondly, a point cloud dataset is only a sample of the reality, reality is represented by a set of points. Going back from a sample of reality to a model is still one of the biggest challenges. Datasets vary in density depending on laser scanner settings, which form a crucial part of the acquisition. Either everything is obtained in high quality or everything is obtained in low quality. No distinction is made between noise data and valuable data (at least not by the laser scanner).

- **How does a successful method look like?**

A successful method looks like below:



Pre-processing is a step which happens no matter the dataset. Then depending on if only parallel bedding planes are to be identified or a fold needs to be identified, different sets of processing steps are performed.

- **How to efficiently extract the required information from large datasets?**

Efficient data extraction is achieved by using your processing tools in a smart way. Filters have to be applied to prevent the whole dense point cloud from being processed by every subroutine. The whole dense point cloud includes a lot of undesired noise slowing down the process. This is why before heavy time intensive computations are done, the point cloud is intensively filtered for dip direction values, intensity values and roughness values.

- **What parameters influence the detectability of the geological features and to what extend?**

There are 3 main parameters which are the main drivers for the detectability of the geological features. Those features are listed below:

- The dataset needs to contain as little artefacts as possible.
- Observed geological layers need to have 3D geometry where layers beddings are visible.
- The distance between the measurement device and the observed geological layers cannot be too large.

If these requirements are not met, results will have a significantly decreased reliability or might even be completely wrong due to non-compatibility of the dataset.

- **How to validate the estimated parameters computed by the computer?**

The validation is done in 2 ways, one way is by comparing the obtained results with available results obtained by previous work. This is done for layer thickness calculations and orientation parameters. The second validation is based on vision, the results are visually inspected in chapter 4.3-4.4 by the maker of this research.

- **How can this method help geologists in assessing geological outcrops?**

This method helps geologists in assessing and slightly improving the results they obtained in the field. The power of this method is that it performs several millions of calculations while a geologist only takes a few measurements. It is a method to validate and potentially also to extract data such as a curvature which a geologist can't or does not observe nor isn't able to quantify.

- ***To what extent can bedding layers and folds in geological outcrops be automatically detected in large 3D datasets?***

Complete automatic detection of bedding layers and folds is not possible yet. The graphical user interface provides an intuitive tool to complete a missing piece of information to smoothen the detection of folds and bedding layers providing a strong base to detect orientation parameters, bedding thickness properties and fold properties. Some user input is still required, mainly for some parameter settings before starting computation. However, this script is more autonomous, faster and more extensive than the previously designed script by Prentice (2017). The only lack of this method compared to its predecessor is that it does not include the separation between limestone and marlstone layers.

7.2. Feedback on goals

In chapter 1, specific goals have been set for this thesis, these goals were the following:

- The extraction capabilities need to be expanded with new geological features which are being extracted.
- The data handling capabilities need to be improved drastically. The previous script took 1 hour to process 2 by 2 meters, and the goal is to process the whole outcrop within a reasonable time span.
- Autonomous processing capabilities need to be increased in comparison to what has already been implemented. This to prevent human errors.

These goals are now evaluated again. This thesis shows that more different features are extracted by the created script, folds are now included in the identification, there are still some cases where folds cannot be extracted but that can be fixed in future research. Identifying bedding layers is not an issue, but it of course has to be mentioned that there are still some commission and omission errors. It has also been proved that this method is very well capable of handling the whole outcrop at once, containing about 50 times more points than before. All calculations are performed within approximately 15 minutes, and lastly the autonomy of the script has greatly been increased. More parameters are now automatically determined by the script, but the method is still semi-automatic because of the remaining parameters and the graphical user interface which is a great tool to improve results.

7.3. Recommendations for future research

This research provides a method to analyze reasonably big sized point cloud datasets with respect to some geological parameters mentioned before. There are however a lot of improvements which could make for some interesting theses. This chapter will elaborate on those potential improvements and other possible recommendations for future work considering geological outcrops.

Faults are also a major feature in geological outcrops. It was originally intended to identify these faults as well, unfortunately this was eventually not possible due to circumstances. Therefore one of the improvements is expanding this script with a method to identify and quantify faults. Figure 52 shows the parameters included when evaluating faults. It is nearly impossible to identify the strike slip component of a fault from just an outcrop, but it is possible to identify and quantify the fault and its normal slip component in a geological outcrop (if the data requirements mentioned in chapter 3.4 are met). One of the features which has been lost during this thesis is division between marlstone and limestone layers thicknesses. This separation should be implemented such that the identification of marlstone and limestone is also automatically performed.

A second recommendation is making the script more autonomous. It has already been improved with respect to autonomy, but it is far from autonomous. Also some pre-processing steps performed by Cloudcompare are done manually. These pre-processing steps can be done using a command line operation. This has to be implemented in future work.

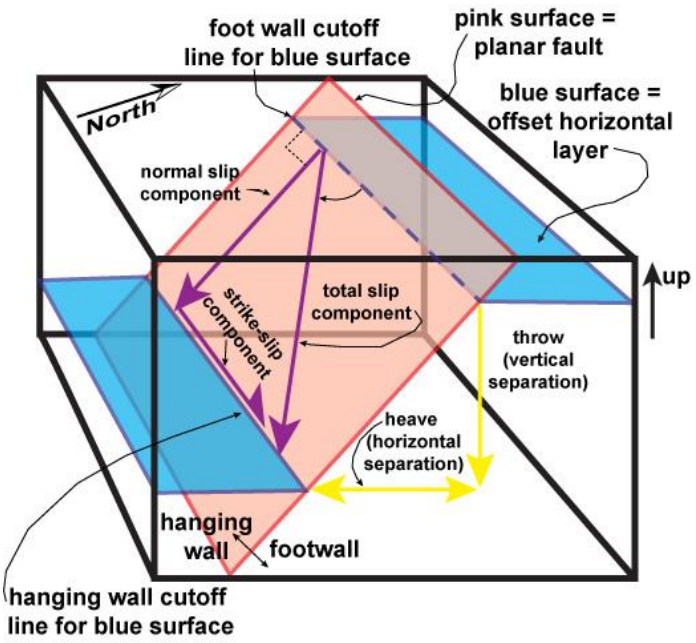


Figure 52: A fault and its parametrization [7].

A far more distant feature such a script should include is producing a full 3D model of the environment. Currently this script is able to process 1 outcrop at the time, the script should be expanded such that it can process multiple point clouds in a row. When correctly georeferenced, these point clouds could then be coupled together in order to create a classification of outcrops and even a 3D model illustrating the layers and their boundaries in the subsurface. The result of such a computation should be comparable to the end result produced by the BSc. students performing the fieldwork in France.

The validation performed in this thesis is not based on error analysis created with the code itself (only partially). The validation is done by comparing the work to work acquired by others as mentioned in chapter 5.1, 5.2 and 5.3. Unfortunately there were no other features to perform a sensible statistical analysis on considering validation. If further analysis is being performed on geological outcrops, a set of criteria has to be established such that validation of results is clearer and easier.

At this instance, validation of results of digital computation of geological outcrops obtained by several researches is not yet established. A comprehensive list for validation would be recommended for further researches.

8 Bibliography

Vosselman, G., & Maas, H. G. (2010). *Arborne and terrestrial Laser Scanning*.

Abellan, A., Derron, M. H., & Jaboyedoff, M. (2016). "Use of 3D Point Clouds in Geohazards" Special Issue: Current Challenges and Future Trends. doi:10.3390/rs8020130

Anders, K., Hammerle, M., Miernik, G., Drews, T., Escalona, A., Townsend, C., & Hofle, B. (2016). 3D Geological outcrop characterization: Automatic detection of 3D planes (azimuth and dip) using LiDAR point clouds. *Annals of the PHotogrammetry, Remote Sensing and Spatial Informaton Sciences,3*(5).

Buckley S.J., Enge H.D., Carlssos C, Howell J.A. (2009 September). Terrestrial Laser Scanning for use in Virtual Outcrop Geology**.**

Corradetti, A., Tavani, S., Russo, M., Cazo, P., & Granado, P. (2017). Quantitative analysis of folds by means of orthorectified photogrammetric 3D models: A case study from Mt. Catria, Northern Apennines, Italy. *The Photogrammetric Record*. doi:10.1111/phor.12212

Humair, F., Abellan, A., Carrea, D., Matasci, B., Epard, J., & Jaboyedoff, M. (2017). Geological layers detection and characterisation using high resolution 3D point clouds: Example of a box-fold in the Swiss Jura Mountains. *European Journal of Remote Sensing,48*(1), 541-568. doi:10.5721/eujrs20154831

Matasci, B., Carrea, D., Abellan, A., Derron, M., Humair, F., Jaboyedoff, M., & Metzger, R. (2015). Geological mapping and fold modeling using Terrestrial Laser Scanning point clouds: Application to the Dents-du-Midi limestone massif (Switzerland). *European Journal of Remote Sensing,48*(1), 569-591. doi:10.5721/eujrs20154832

Schober, A., & Exner, U. (2011). 3D structural modelling of an outcrop-scale fold train using photogrammetry and GPS mapping. *Austrian Journal of Earth Sciences,104*(2), 73-79.

Silva, R. M., Veronez, M. R., Gonzaga, L. J., Tognoli, F. M., Souza, M. K., & Inocencio, L. C. (2015). 3-D Reconstruction Of Digital Outcrop Model Based On Multiple View Images And Terrestrial Laser Scanning. 245-253.

Walstra J., Dixon N., & Chandler J. (2007). Historical aerial photographs for landslide assessment: Two case histories. *Quarterly Journal of Engineering Geology and Hydrogeology,40*(4), 315-332. doi:10.1144/1470-9236/07-011

Westoby M.J., Brasington J., Glasser N.F., Hambrey M.J., Reynolds J.M. (2012). 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology 179. doi:* https://doi.org/10.1016/j.geomorph.2012.08.021

Torr, P.H.S., & Zisserman, A. (1996). MLESAC: A new robust estimator with application to estimating image geometry.

Bisschop, T. (2017, October 23). Virtual Outcrop Analysis using Lime. BSc. Thesis TU Delft.

Brittanica, T.E. (2014, February 20). Stratification. Retrieved June 29, 2018, from https://www.britannica.com/science/stratification-geology

E. (2008, January 11). Milankovitch Cycles and Stratigraphy. Retrieved from http://dynamic-earth.blogspot.nl/2008/02/milankovitch-cycles-and-stratigraphy.html

Fisher, B. (2002, May 6). The RANSAC (Random Sample Consensus) Algorithm. Retrieved from http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FISHER/RANSAC/

Grippo, A. (2016, May 17). Sea level change. Retrieved from http://homepage.smc.edu/grippo_alessandro/gss7.html

US Department of Commerce, & National Oceanic and Atmospheric Administration. (2018, June 25). What is LIDAR. Retrieved from https://oceanservice.noaa.gov/facts/lidar.html

[1] Geology In. (n.d.). Principle of original horizontality. Retrieved from http://www.geologyin.com/2014/03/principle-of-original-horizontality.html.

[2] Sedimentary Structures. (n.d.). Retrieved from http://www.indiana.edu/~geol105/images/gaia_chapter_5/sedimentary_structures.htm

[3] Milankovitch Cycles and Glaciation. (n.d.). Retrieved June 29, 2018, from http://www.indiana.edu/~geol105/images/gaia_chapter_4/milankovitch.htm

[4] (n.d.). Retrieved June 29, 2018, from http://www.photogrammetry.com/

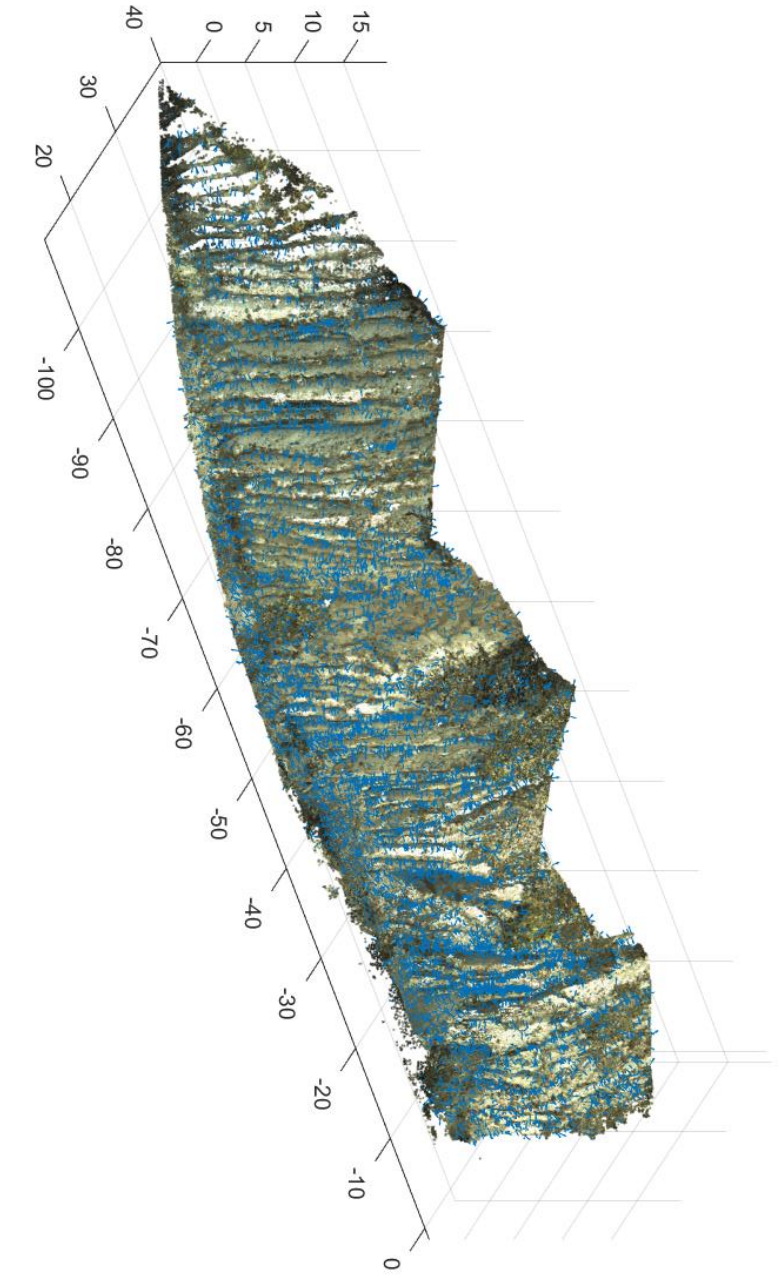
[5] Renishaw. (n.d.). Application note: Optical encoders and LiDAR scanning. Retrieved June 29, 2018, from http://www.renishaw.com/en/optical-encoders-and-lidar-scanning--39244

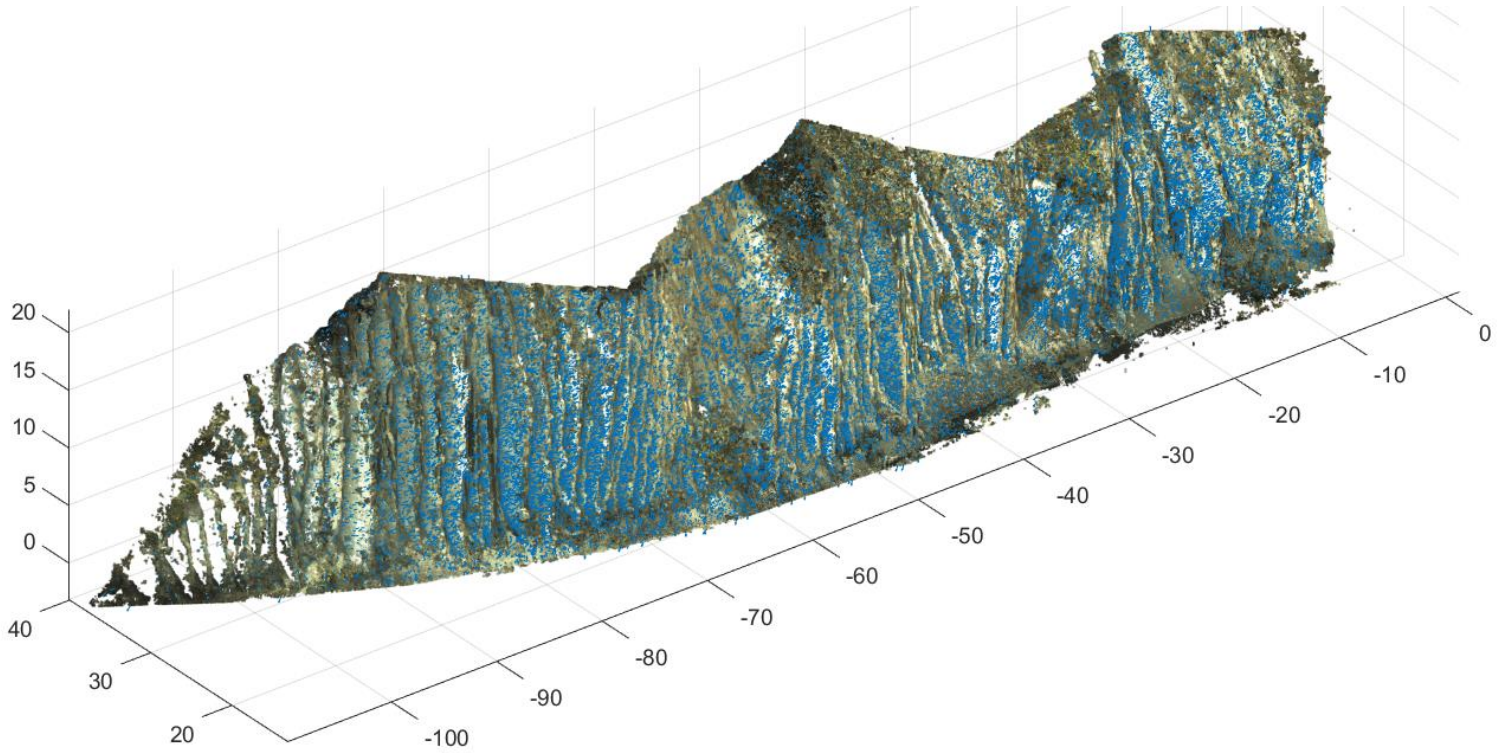
[6] Detail of curved beddings. (n.d.). Retrieved from https://dissolve.com/stock-photo/Detail-curved-bedding-planes-rock-water-edge-County-royalty-free-image/101-D869-7-856

[7] Descriptive structural geology of faults. (n.d.). Retrieved from http://maps.unomaha.edu/Maher/GEOL3300/week2/fault.html

9 Appendices

Appendix 1




20
15
10
5
0
40
30
20
-100
-90
-80
-70
-60
-50
-40
-30
-20
-10
0

Appendix II

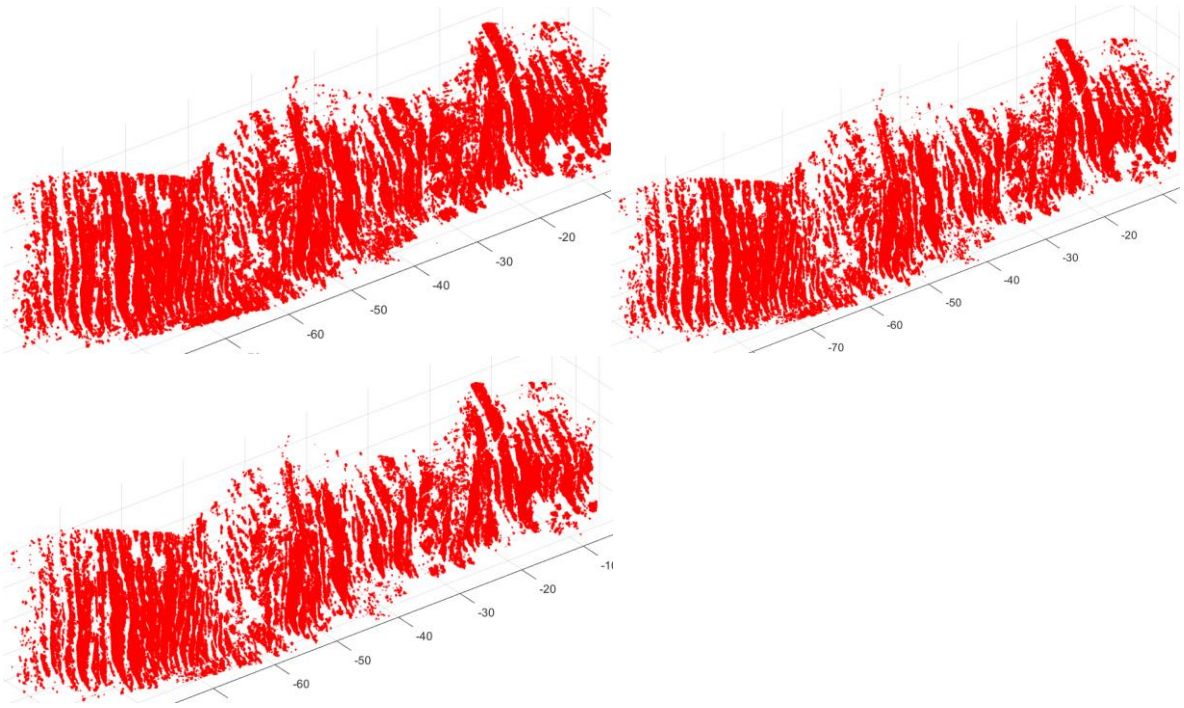


Figure1: Threshold denoise. Thresholds of 0.1m, 0.024m and 0.0001m.

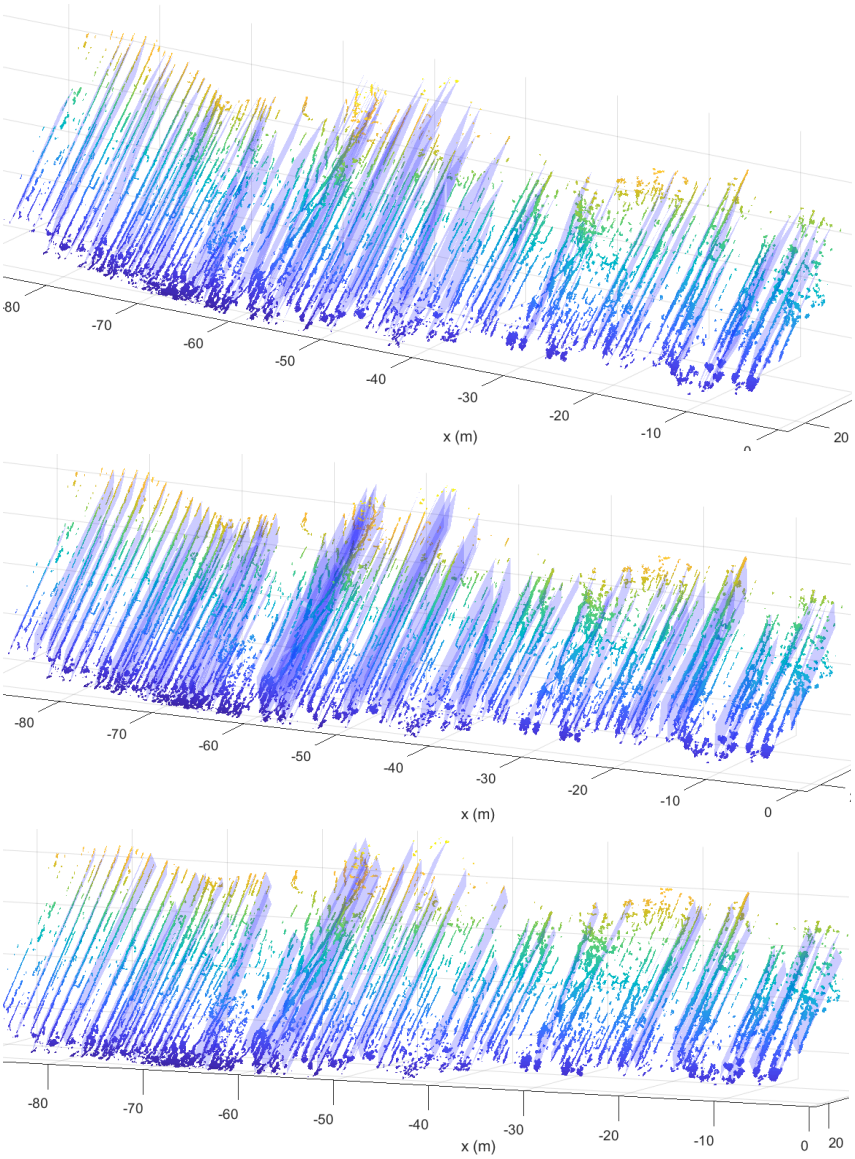


Figure 2: ***Minimum points per plane. 1000 points, 2000 points and 5000 points.***

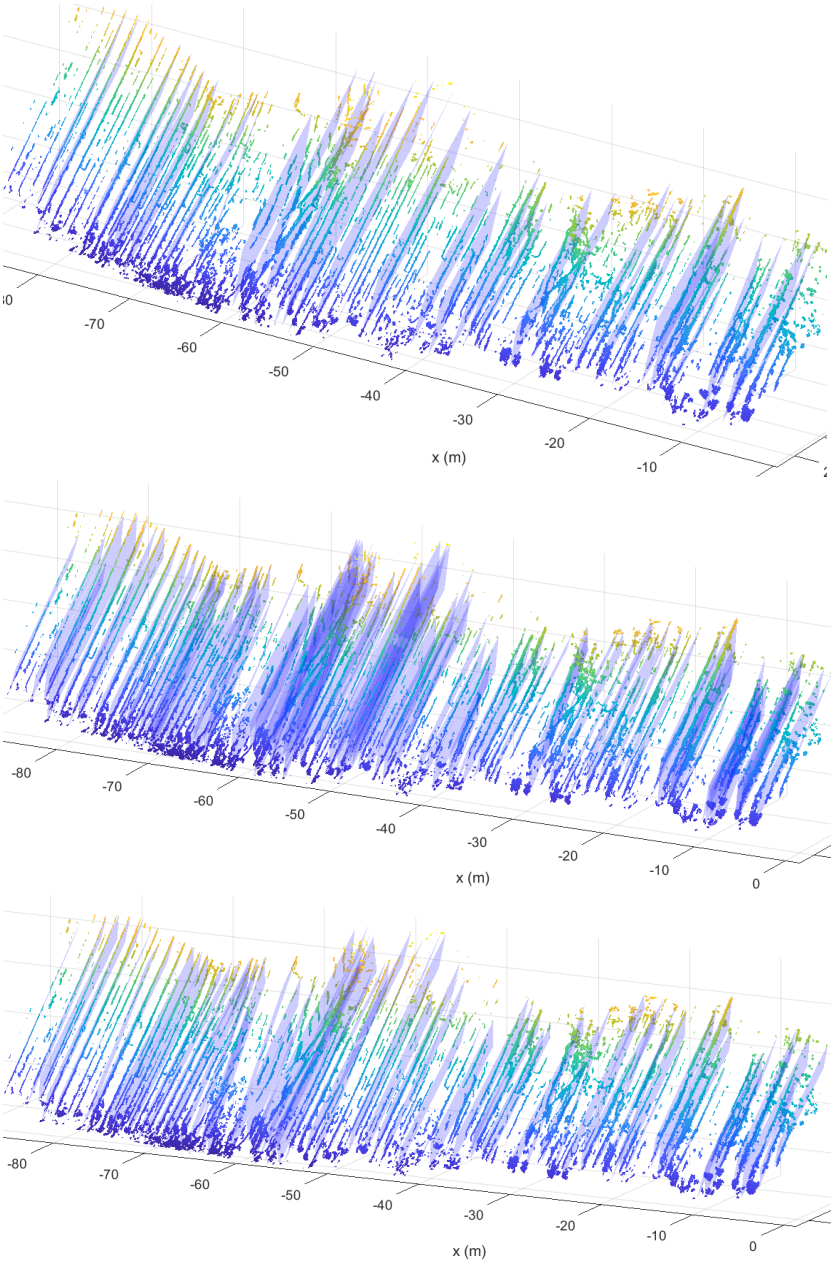


Figure 3: ***Maximum distance plane fitting. 7 cm, 15 cm 25 cm.***

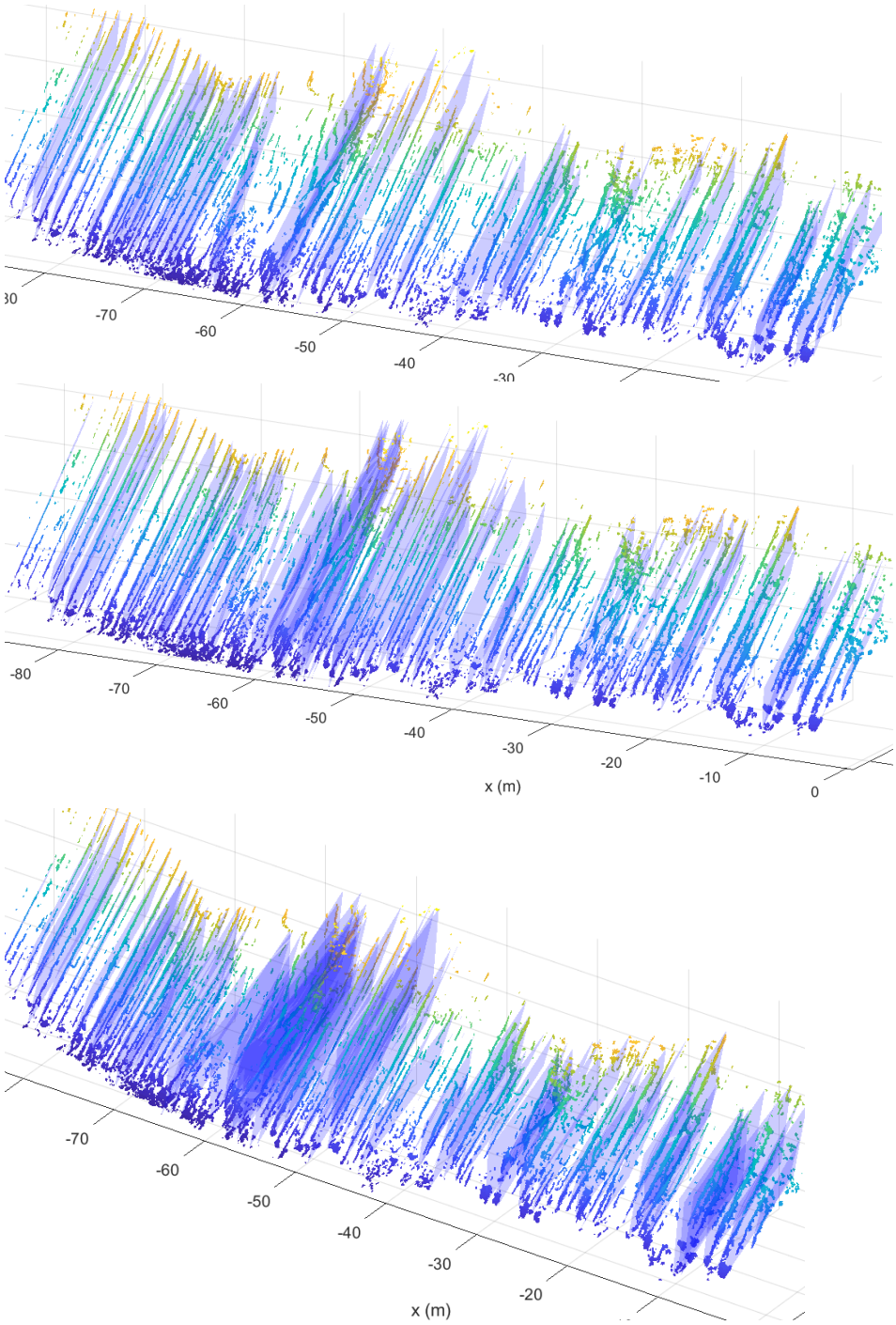


Figure 4: ***Maximum distance plane fitting. 1 degree, 5 degrees 10 degrees.***

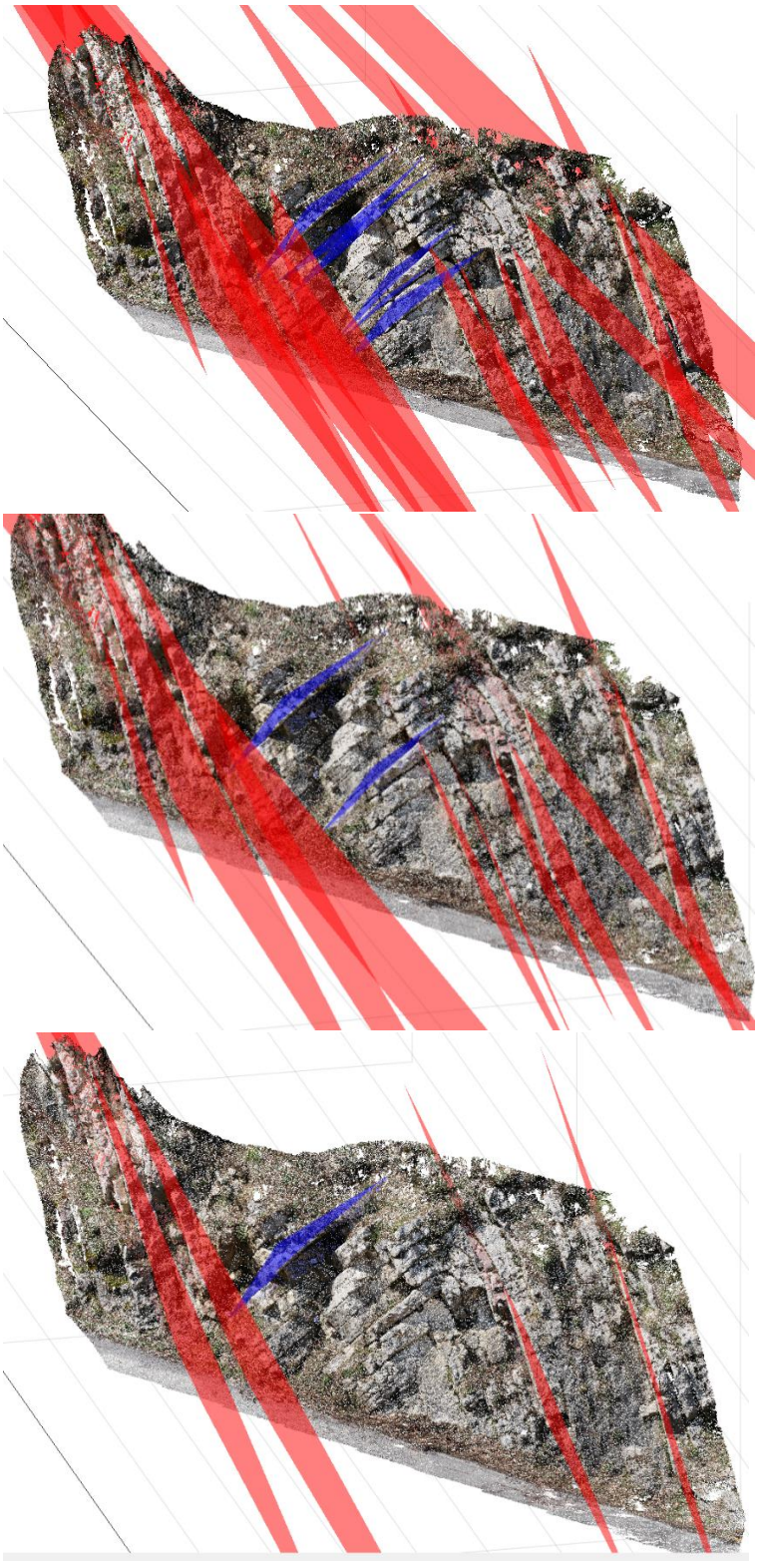


Figure 5: Minimum points per plane. 500 points, 1000 points and 2000 points.

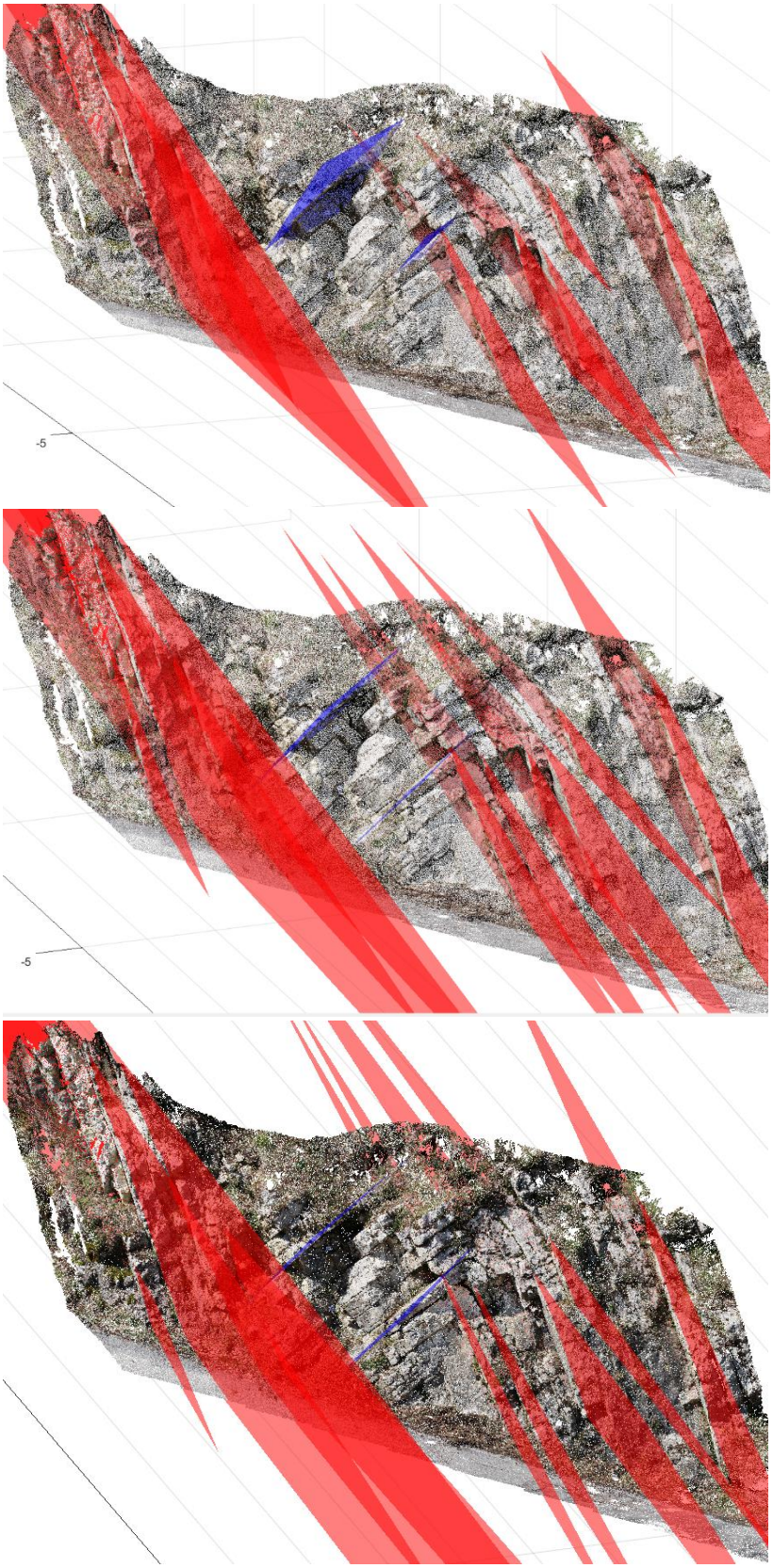


Figure 6: Maximum distance plane fitting. 7 cm, 15 cm and 25 cm.

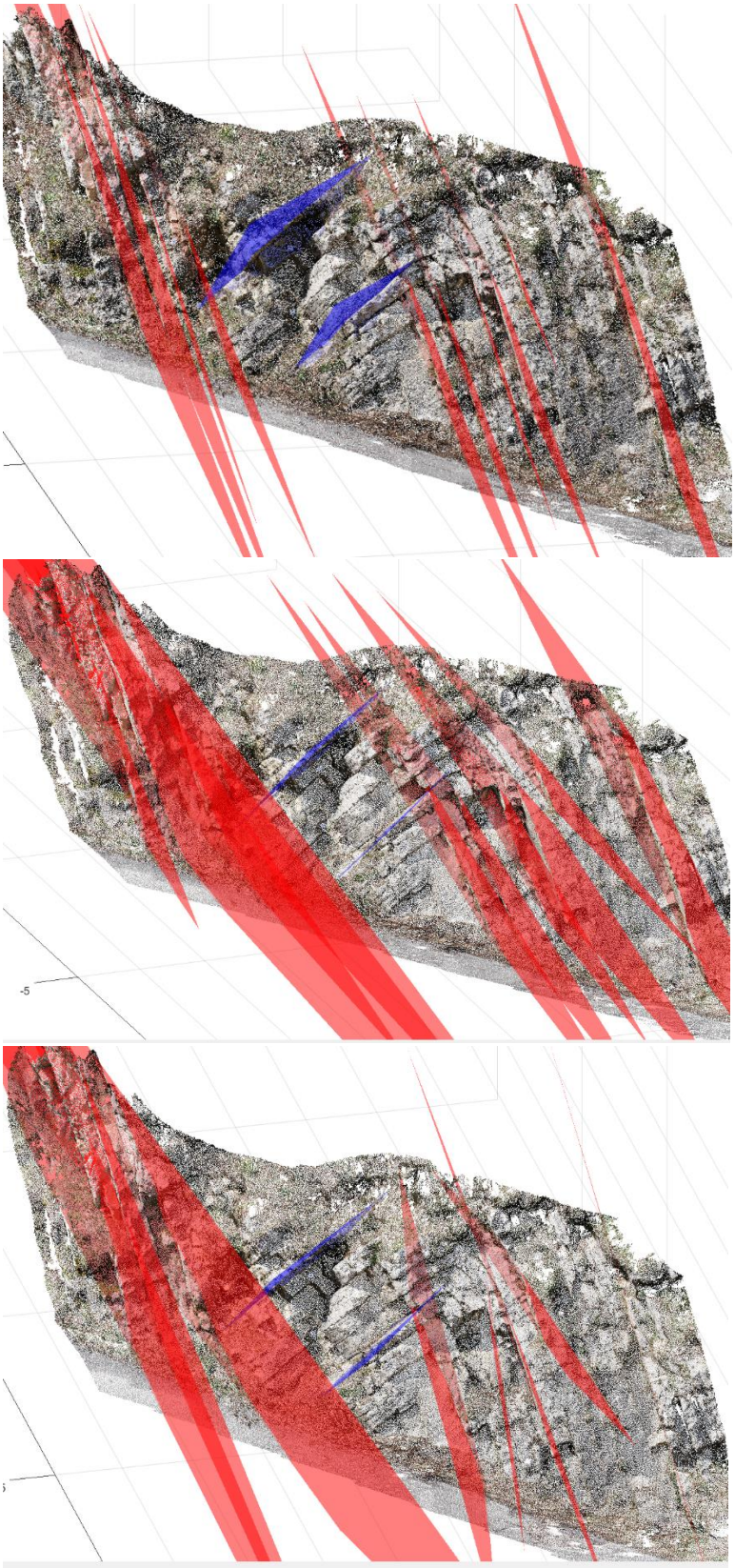


Figure 7: Angular difference plane fitting. 1 degree, 5 degrees and 10 degrees.

Appendix III

- *Cloud subsampling based on Octree.*
  The cloud is subdivided into a simpler version of itself in terms of space. The nearest point to the octree cell center is kept. The higher the Octree level, the smaller the cells. The octree level is an equal subdivision of the space into boxes, every increase in octree level subdivides every box into 4 equal sized smaller boxes.
- *Random Cloud subsampling.*
  The user defines X amount of points which the program cannot exceed. Cloudcompare then randomly picks X points from the point cloud, not making any distinction with respect to geometry.
- *Scalar field subsampling.*
  The scalar field subsampling provides the option to also take into account the scalar values of the points. One can now also subsample based on planarity.
- *Filter by value.*
  The filter by value option within Cloudcompare provides the option to directly filter out scalar field values exceeding a certain range. This can help filtering out points with a certain intensity or orientation, serving as a filter.
- *Label connected components.*
  The label connected components tool in Cloudcompare is a segmentation tool which divides the point cloud into partitions which presumably belong together. An Octree level defines the gap between the cells, the larger the Octree level the smaller the gap. A minimal amount of points can be set for every segment, with which segment sizes can be determined.
- *Range and angle correction.*
  Range and incidence angle of the laser beam can be used in order to correct intensity measurements. A Quadratic correction is applied for the range and a cosine correction is applied for the incidence angle.
- *Voxtree.*
  Method which distances between points in order to find objects like trees. Might be usable to filter out bushes
- *Region growing based segmentation.*
  Point clouds can be separated by region based segmentation. This segmentation start with a seed point (or seed points) and tends to grow from there until the edge of the segment is reached. Then a new seed point is chosen, and so on. Confirmed by
- *MSAC*
  M-estimator Sample and Consensus (MSAC) is an alternative method of fitting planes through a point cloud. The difference with RANSAC is that points with a distance smaller and larger than the threshold do not get a penalty of 0 resp. C (constant), but a penalty based on the distance squared. It is 5-10% better than RANSAC with respect to speed and accuracy (Torr. 1996).
- *Hough Transform.*
  The Hough Transform (HT) is used for segmentation of point clouds. It does not fit random shapes to the point clouds like in RANSAC, but it searches for random shapes in the point clouds. In that sense it belongs to the same family as RANSAC

- *Segment-based filtering.*
  Most filtering has been performed on points, this method applies filtering on segments. It combines filtering with segmentation and actually filters out segments which have undesired properties. “Airborne and terrestrial Laser Scanning”, Edited by George Vosselman and Hand-Gerd Maas.
- *KNNSearch matlab*
  The KNNSearch function in Matlab searches for the K nearest values using data. It can be used for intensity filtering.
- *K-means clustering.*
  K-means clustering helps with defining what points belong to the same species. When normals of the points are computed, clustering helps defining groups of normals which belong together. This way, rocks surface layers and bushes can be separated.
- *Stereoplot.*
  A stereoplot can be created and used in order to visually address the problem of clustering. The stereoplots can help in identifying clusters and eventually also identify folds if used correctly.
- *Triangular Irregular Network.*
  The Triangular Irregular Network methods (called TIN) is also used to extract 3D meshes form point clouds. It can be used in searching irregularities (maybe faults). It is widely used in meshing geographic surfaces [3]. An example of this TIN methods is the (2.5D) Delaunay Triangulation.
- *KD-tree.*
  The KD-tree is used for hierarchical data structures for nearest neighbor searching in 3D points clouds.
- *Irregularities.*
  There is a method described in which provides the opportunity to search for similarities between layers and objects. It can be used in order to find for example faults. A fault is a discontinuity of a certain amount of layers. A fault has an orientation, and along this orientation similarities can be observed (providing proof for a fault being present).